# Circulating Tumour DNA in Localised Urological Cancers


CANCER RESEARCH UK | CAMBRIDGE INSTITUTE

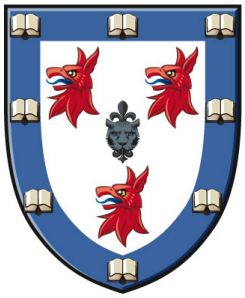


UNIVERSITY OF CAMBRIDGE

**THIS DISSERTATION IS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY BY**

**DR. KEVAL MAHENDRA PATEL**

**SUPERVISED BY: DR. NITZAN ROSENFELD**



**HOMERTON COLLEGE**

**MARCH 2017**

# ABSTRACT

There is a need for informative biomarkers in localised urological cancers. At present, no method can accurately distinguish between indolent and aggressive prostate cancers, and men often require repeated biopsies. Patients with muscle invasive bladder cancer (MIBC) undergo neo-adjuvant chemotherapy (NAC) to improve survival. However, many do not respond to NAC unnecessarily delaying definitive treatment. Cell-free mutant DNA (mutDNA) analysis represents an opportunity for non-invasive monitoring of cancer through tumour genome analysis. MutDNA derived from plasma can monitor tumour burden. There is emerging evidence that mutDNA can identify mutations from multiple clones and is abundant in adjacent body fluids. This work explores the utility of plasma and urinary mutDNA in localised prostate and bladder cancers.

This thesis describes the optimisation of urinary mutDNA analysis by assessing urinary DNA processing and extraction methods using healthy volunteer and bladder cancer patient urine samples. Primer panels were designed and validated to target frequently mutated regions in prostate and bladder cancers, as well as for analysis of patient-specific mutations. Sequencing-based methods and digital Polymerase Chain Reaction (dPCR) were employed to analyse clinical samples including plasma and urine, to detect and quantify mutDNA. Molecular and clinical data were integrated to explore potential areas of application of mutDNA analysis.

For bladder cancer, mutDNA was analysed from liquid-biopsy samples including plasma, cell pellets from urine (UCP) and urine supernatant (USN) from multiple time-points of 17 MIBC patients undergoing NAC. I showed that mutDNA was more frequently detected and was present at higher AFs in urine compared to plasma samples. Of potential clinical relevance, I showed that the presence of mutDNA after starting NAC was associated with disease recurrence. This original contribution to knowledge could offer patients an opportunity to expedite surgical resection in a timely manner, if corroborated in large-scale trials.

For prostate cancer, a *TP53* specific panel was applied to men with metastatic disease, to demonstrate that clones containing *TP53* mutations, which are dominant in at the metastatic stage were present in historical prostatectomy samples taken when then patient was believed to have localised disease only. Furthermore, I showed that these *TP53* mutations could be detected at the localised stage of disease. To investigate the ability of mutDNA detection private clonal mutations I developed a method for higher sensitivity

analysis (MRD-Seq). This was applied to a clinical cohort of 2 men with multi-focal localised prostate cancer to demonstrate the though the overall levels of mutDNA is low, private clonal mutations may be detectable. Taken together, these original contributions to knowledge could allow for less invasive surveillance of men with low risk prostate cancer and warrants further investigation.

In this thesis, I used a range of molecular methods were applied to small cohorts of clinical samples from patients with urological malignancies, in an exploratory analysis. The molecular data was analysed in conjunction with clinical information to draw hypotheses on the biology and natural history of these cancer, and to suggest possible utility of mutDNA analysis in their clinical management. Some of the findings suggest areas of potential utility, which merit further validation or investigation in larger cohorts or clinical studies.

# ACKNOWLEDGEMENTS

I would firstly like to thank my supervisor, Dr. Nitzan Rosenfeld, and my post-doctoral supervisors, Dr. Tim Forshew, Dr. Dana Tsui, Dr. Charlie Massie, Dr. Florent Mouliere and Dr. Chris Smith for their invaluable advice and support. I would also like to thank all members of Rosenfeld research group and collaborating research groups who took time out of their busy schedules to guide and help me. Without their help, these investigations would not have been possible.

I am very grateful for the generous support of my sponsors, the Royal College of Surgeons of England, the Addenbrooke's Charitable Trust and Cancer Research UK Cambridge Centre.

Finally, I would like to thank my wife and family without whose support, I would not have been able to complete this work.


# STATEMENT OF ORIGINALITY

I declare that the following ideas and work is my own, except where referenced otherwise. Any collaborative work is specifically declared in relevant text.

This work has not been submitted for a degree or diploma at the University of Cambridge or any other institution before and is not concurrently being submitted for any such degree.

This doctoral thesis does not exceed the prescribed work limit of 60,000 words set by the Clinical Medicine and Clinical Veterinary Medicine Degree Committee.

# TABLE OF CONTENTS

# FIGURES

# TABLES

# ABBREVIATIONS

| | |
|---|---|
| **AF** | **Allele Fraction** |
| **BC** | **Bladder Cancer** |
| **BCG** | **Bacillus Calmettee Guerin** |
| **BLAT** | **BLAST like Alignment Tool** |
| **bp** | **base pair** |
| **BR** | **Biochemical Recurrence** |
| **BUF** | **Buffy coat** |
| **BWA** | **Burrows-Wheeler Alignment** |
| **CA15-3** | **Cancer Antigen 15-3** |
| **CAPRA** | **Cancer of the Prostate Risk Assessment** |
| **cfDNA** | **Cell-Free DeoxyriboNucleic Acids** |
| **CI** | **Confidence Interval** |
| **CIS** | **Carcinoma In-Situ** |
| **CNAs** | **Copy Number Aberrations** |
| **COSMIC** | **Catalogue Of Somatic Mutations In Cancer** |
| **CR** | **Chemoradiotherapy** |
| **CRPC** | **Castrate Resistant Prostate Cancer** |
| **CTCs** | **Circulating Tumour Cells** |
| **ctDNA** | **Circulating Tumour DeoxyriboNucleic Acids** |
| **DNA** | **DeoxyriboNucleic Acids** |
| **dPCR** | **Digital Polymerase Chain Reaction** |
| **ECDF** | **Empirical Cumulative Distribution Function** |
| **EDTA** | **Ethylenediaminetetraacetic acid** |
| **EORTC** | **European Organization for Research & Treatment of Cancer** |
| **FFPE** | **Formalin Fixed Paraffin Embedded** |
| **GC content** | **guanine-cytosine content** |
| **GE copies** | **Genomic Equivalent copies** |
| **HVP** | **Healthy Volunteer Plasma** |
| **ICGC** | **International Cancer Genome Consortium** |
| **Indel** | **Insertions or Deletions** |
| **ISUP** | **International Society of Urological Pathology** |
| **LND** | **Lymph Node Dissection** |
| **mAF** | **Mutant Allele Fraction** |
| **MCDNA** | **Male Genomic (Control) DNA from healthy volunteers** |
| **MIBC** | **Muscle Invasive Bladder Cancer** |
| **MRD-Seq** | **Multiple Replicate Dilution Sequencing** |
| **MRI** | **Magnetic Resonance Imaging** |
| **mutDNA** | **Mutant DNA in body fluids including plasma and urine (so called as we analysed both cell-free DNA and cellular DNA)** |
| **NA** | **Nucleic Acid** |
| **NAC** | **Neo-Adjuvant Chemotherapy** |
| **NCCN** | **National Comprehensive Cancer Network** |
| **NGS** | **Next Generation Sequencing** |

| | |
|---|---|
| **NICE** | **National Institute for Health and Care Excellence** |
| **NKI** | **National Cancer Institute** |
| **NMIBC** | **Non-Muscle Invasive Bladder Cancer** |
| **OS** | **Overall Survival** |
| **PC** | **Prostate Cancer** |
| **pCR** | **pathological complete response** |
| **PCR** | **Polymerase Chain Reaction** |
| **PCSM** | **Prostate Cancer Specific Mortality** |
| **PIN** | **Prostatic Intra-epithelial Neoplasia** |
| **PLS** | **Plasma** |
| **pPD** | **pathological progressive disease** |
| **pPR** | **pathological partial response** |
| **PSA** | **Prostate Specific Antigen** |
| **pSD** | **pathological stable disease** |
| **qPCR** | **Quantitative PCR** |
| **Rad** | **Radiotherapy** |
| **RC** | **Radical Cystectomy** |
| **RNA** | **RiboNucleic Acid** |
| **SCC** | **Squamous Cell Carcinoma** |
| **SNP** | **Single Nucleotide Polymorphism** |
| **SNV** | **Single Nucleotide Variant** |
| **SOP** | **Standard Operating Procedure** |
| **sWGS** | **Shallow Whole Genome Sequencing** |
| **TAm-Seq** | **Tagged Amplicon deep Sequencing** |
| **TRUS** | **TransRectal Ultrasound Biopsy** |
| **TUR** | **Tumour** |
| **U.K.** | **United Kingdom** |
| **UCC** | **Urothelial Cell Carcinoma** |
| **UCP** | **Urine Cell Pellet** |
| **UCSC** | **University of California, Santa Cruz** |
| **USA** | **United State of America** |
| **USN** | **Urine Supernatant** |
| **WGA** | **Whole Genome Amplified** |
| **WGS** | **Whole Genome Sequencing** |

# CHAPTER 1: INTRODUCTION

*The ideas presented in this chapter were first described in my first year report and parts were published in Clinical Biochemistry and Journal of Clinical Urology. The text is therefore to a large extent excerpted from my first year report and the related publication, but has been updated to incorporate novel ideas, findings and up-to-date references.*

> **K.M. Patel**, D.W.Y. Tsui. The translational potential of circulating tumour DNA in oncology. Clinical Biochemistry. October 2015. 48:15, 957-961. ISSN 0009-9120, http://dx.doi.org/10.1016/j.clinbiochem.2015.04.005.
>
> **K.M. Patel**, V.J. Gnanapragasam, Novel concepts for risk stratification in prostate cancer. Journal of Clinical Urology. December 2016. 9(2), 18-23, doi:10.1177/2051415816673502.

## 1.1 Circulating Tumour DNA

In 1948, Mandel *et al.* were the first to describe cell free nucleic acids circulating freely in the blood (Mandel and Metais, 1948). It was not until 1977, when Leon *et al.* compared serum DNA levels (later termed circulating free DNA or cfDNA) of 173 cancer patients and 55 healthy volunteers, did the realisation that circulating DNA could be an important marker of disease burden occur (Leon *et al.*, 1977).

Unfortunately, measurement of cfDNA alone has a great overlap between healthy volunteers and cancer patients and is therefore not sensitive (Jung *et al.*, 2010; Zhang *et al.*, 2010; Gonzalez-Masia *et al.*, 2013; van der Vaart and Pretorius, 2010). It is also non-specific, being raised in numerous conditions ranging from pregnancy, autoimmune conditions, trauma, pancreatitis, cancer and myocardial infarction (Bartoloni *et al.*, 2011; Chen *et al.*, 2011; Antonatos *et al.*, 2006; Ha *et al.*, 2011; Lui and Dennis, 2002).

However, all tumours contain genetic alterations and, rare DNA fragments in blood contain tumour-specific mutations (circulating Tumour DNA or ctDNA). By designing PCR assays for known mutation hotspots in pancreatic cancer, Sorenson *et al.* were able to demonstrate the presence of ctDNA in blood almost twenty years ago (Sorenson *et al.*, 1994). A more personalised approach was utilised by Diehl *et al.* who used Sanger sequencing to identify point mutations in the tumour of each patient, then by designing mutation specific fluorescent probes, were able to detect mutant copies of DNA in the plasma of colorectal patients (Diehl *et al.*, 2005). In a subsequent study evaluating 162

plasma samples from 18 patients, Diehl *et al.* demonstrated that levels of ctDNA correlate well with tumour burden and was more sensitive than serum measurements of Carcino-Embryonic Antigen, the current standard biomarker, for monitoring disease burden (Diehl *et al.*, 2008).

Several groups demonstrated the presence of tumour specific point and rearrangement mutations in the plasma of patients with different cancer types. In each case, the initial step would be to determine the mutational profile of the tumour tissue or to use known mutation locations to then design tumour specific assays to track ctDNA levels in the blood (Leary *et al.*, 2010; McBride *et al.*, 2010; Chen *et al.*, 1999; Yung *et al.*, 2009; Otsuka *et al.*, 2004; Diaz *et al.*, 2012; Dawson *et al.*, 2013).

### 1.1.1 Monitoring Tumour Evolution With ctDNA

The above techniques rely on having prior knowledge of the mutational profile of the tumour. This is achieved as either; there are few known tumour-specific mutations to target or, through tumour re-sequencing of patient derived tissue. However in 2012, Forshew *et al.* demonstrated that novel mutations can be detected directly through plasma re-sequencing (Forshew *et al.*, 2012). For a woman with ovarian cancer, re-sequencing of tumour tissue from right oophorectomy specimen, taken at the time of de-bulking surgery, had shown *TP53* mutation. ctDNA levels of *TP53* were tracked using TAm-Seq. As the patient progressed, TAm-Seq of plasma samples showed the presence of an *EGFR* mutation, not initially found in the oophorectomy specimen. On further investigation, the same *EGFR* mutation was present at low frequency in samples taken from the Omentum at the time of de-bulking surgery. It is likely that this clone became more dominant as chemotherapy regimes contained other clones.

In 2013, the Rosenfeld group, in collaboration with the group of Professor Carlos Caldas, investigated 30 women with metastatic breast cancer. Tumour tissue mutations were identified using next generation re-sequencing and subsequently, TAm-Seq was used to successfully detect ctDNA in 29 out of the 30 women. Computerised Tomograms were compared with levels of ctDNA, Cancer Antigen 15-3 (CA15-3) and Circulating Tumour Cells (CTCs) taken with from serial samples. Overall, ctDNA was able to detect changes in tumour burden faster and with greater sensitivity than either the standard biomarker (CA15-3) or CTCs (Dawson *et al.*, 2013).

By following 6 patients with metastatic tumours, Murtaza *et al.* were able to demonstrate that direct sequencing of the plasma can be used to track tumour evolution in response to therapy and to detect novel mutations. This was achieved by selecting 6 patients that had high fractions of mutant:wild type alleles in their plasma circulating DNA. Further analysis was conducted by generating barcoded and pooled libraries using the commercially ThruPLEX-FD kit (Rubicon genomics), followed by library enrichment using the TruSeq Exome Enrichment Kit (Illumina). Subsequent re-sequencing and variant analysis revealed that plasma ctDNA may harbour additional mutations, when compared to biopsy material. These additional mutations are likely to come from tumour clones underrepresented in the original biopsy (Murtaza *et al.*, 2013).

### 1.1.2 Detecting ctDNA in Low Volume Disease

Several approaches have been used to detect ctDNA, each with differing analytical sensitivities and associated costs. On the one extreme, dPCR can be used to target known mutations or mutation hotspots, however the number of genomic locations targetable are low due to the inability to significantly multiplex reactions. However, the analytical sensitivity is high, reported at 1 in 13,000 mutant:wild-types (Milbury *et al.*, 2014) and the cost per sample, relatively low. On the other extreme are WGS approaches, which prepare sequencing libraries directly from cfDNA. WGS approaches are therefore unbiased, assessing all genomic locations, and can detect previously unknown mutations. However, the cost is high due to the sequencing needed to attain even modest coverage. Subsequently, the analytical sensitivity is limited to 5-10 in 100 mutant:wild-types molecules (Heitzer *et al.*). Amplicon based and capture based sequencing form the 2 intermediate approaches used and accommodate reduced breadth of genomic interrogation for increased sequencing reduced sequencing cost and improved coverage. Importantly, as the cost of sequencing continues to drop, the feasibility of attaining higher coverage increases and hence improve analytical sensitivity.

As the sensitivity of ctDNA analysis improves, the unbiased detection of ctDNA at lower levels will become more feasible. Currently many cancer patients still present too late for curative therapies and our best chance of improving cancer mortality rates lie in the early detection of cancer (Vogelstein *et al.*, 2013). However, at present published studies have largely focused on ctDNA analysis in the advanced cancer setting, where levels of mutant:wild type allele fractions is around 10-50% (Murtaza *et al.*, 2013; Forshew *et al.*, 2012; Diehl *et al.*, 2005). For earlier stages, the mutant:wild type allele fractions are likely

to be lower. Despite the challenges at the onset of my investigations, several studies used dPCR or amplicon sequencing strategies to detect ctDNA in several, non-urological, cancers from the plasma of patients at an early stage of disease.

In Diehl *et al.*'s seminal work, mutant:wild type allele frequencies ranging from 0.001-0.12% were detected in patients with localised colorectal cancer. Of note, ctDNA was also detected from plasma samples in their control group of patients with adenoma, with mutant:wild type allele frequencies of 0.001–0.02% (Diehl *et al.*, 2005). The ability to detect ctDNA in patients with adenomatous colorectal polyps has important implications. Colorectal adenomas are precursors to adenocarcinoma and as such, often require further treatment. That ctDNA was detected at this pre-cancerous stage suggests a potential role of ctDNA in early diagnosis or even for population screening.

In more recent work, Bettegowda *et al.* reported a range of 0 (not-detected) to 135,000 mutant fragments/5ml of blood in 223 patients with 12 tumour types, all with localised disease (Bettegowda *et al.*, 2014). By detecting mutations in tumour tissue and subsequently searching for these in matched plasma samples, they detected ctDNA in 55% (122/223) of patients (Bettegowda *et al.*, 2014). Furthermore, by using a digital droplet PCR (Taly *et al.*, 2012) assay, Beaver *et al.* were able to detect PIK3CA mutations in the pre-operative plasma samples of 14/15 patients with low volume breast cancer and who had PIK3CA mutations detected in their tumour samples (Beaver *et al.*, 2014). For 10 patients, blood samples were collected post-operatively and 5 had persistently detectable ctDNA. Indeed, one of these patients went on to have cancer recurrence. The ability of ctDNA to detect and track tumour heterogeneity, coupled with the ability to detect low volume disease could be of vital importance in urological cancers where currently tumour heterogeneity leads to much clinical uncertainty (as outlined below in section 1.2.3).

### 1.1.3 cfDNA in other bodily fluids

Much ctDNA research to-date focuses on plasma and serum derived samples. However, frequent blood sampling in patients who are already prone to anaemia of chronic disease is not ideal. Indeed, ethical committees often limit the amount of blood that can be drawn for study purposes. Interestingly, tumour-specific nucleic acids have been detected in other bodily fluids, including stool (Vogelstein and Kinzler, 1999), urine (Millholland, 2012), saliva (Li *et al.*, 2004), cerebrospinal fluid (Pan *et al.*, 2015) and pleural fluid (Soh *et al.*, 2006). To refer to mutant DNA extracted from bodily fluids, including plasma and urine, I use the term mutDNA, as the DNA of interest arises from circulating and non-circulating

fluids and can be cellular or cell-free. Particular fluids may concentrate mutDNA from regional drainage and may facilitate low volume disease mutDNA detection. For example, mutated DNA fragments from localised oral cancers are hypothesized to be more concentrated in saliva than in the circulation. Additionally, many bodily fluids are freely available and found in abundance, particularly urine. This may allow more frequent and more voluminous sampling aiding low volume disease detection.

In likeness to ctDNA found in plasma, the presence of mutDNA in the urine has long been established (Sidransky *et al.*, 1991). However, the DNA detected by Sidransky *et al.* was cellular in origin. For bladder cancer, cellular diagnostic approaches (urinary cytology) are widely used clinically for monitoring. Despite this, abnormal cells can be absent in more than 50% of cases, particularly for low-grade urothelial cancers (Simon *et al.*, 2003; Babjuk *et al.*, 2014).

In 1999, Zhang *et al.* demonstrated the presence of *SRY* gene (found on the Y chromosome) in the urine supernatant of female recipients of male donor renal transplants (Zhang *et al.*, 1999). By 2000, Botezatu *et al.* demonstrated that some genomic DNA escaped into the urine through renal filtration (Botezatu *et al.*, 2000). They demonstrated this by subcutaneously injecting mice with radiolabelled human DNA and collecting urine over the next 3 days. Collected urine was determined to contain human cfDNA through its radioactivity and PCR amplification of human specific *Alu* repeat sequences Subsequently, by performing nested PCR to detect highly repetitive regions located on the Y chromosome, Botezatu *et al.* demonstrated the presence of male DNA in the urine of pregnant women carrying male foetuses (Botezatu *et al.*, 2000).

In 2004, Su *et al.* demonstrated the presence of mutated DNA in the urine of colorectal carcinoma patients. Using Restriction Enriched PCR with 46 base primer pairs targeting a known mutation in *KRAS* codon 12, they were able to selectively amplify mutated DNA in the fractionated urine of colorectal cancer patients.

At the onset of my investigations, next generation sequencing (NGS) approaches had been used to detect mutDNA derived from the urine of cancer patients. In 2012, Millholland *et al.* targeted the known hotspots in *FGFR3* exons 7, 10 and 15 and detected *FGFR3* mutations in urinary DNA (Millholland, 2012). They reported a concordance of 91% by detecting mutant urinary cfDNA in 10 out of 11 patients with *FGFR3* mutations in their primary tumours. In a separate group, the investigators assessed the urine of 43 known patients with bladder cancer and found mutated *FGFR3* in 24 of 43 cases (Millholland,

2012). As the mutation status of the primary tumour was not reported, no comment can be made on the concordance of their urinary assay with tumour mutations, though it appears proportional to the prevalence of *FGFR3* mutations in bladder cancer populations detected in other studies (Lamy *et al.*, 2006; Sjodahl *et al.*, 2011).

Similar findings were also presented in previous studies relying on molecular assays to analyse *FGFR3* mutations (van Kessel *et al.*, 2012; Zuiverloon *et al.*, 2010). The ability of the assays to detect bladder cancer mutations as a whole was, however, low due to the sole focus on *FGFR3*, despite some studies selecting patients who had *FGFR3* aberrations in their initial tumour sample.

The findings of Botezatu and Su demonstrate that renally filtered cfDNA is available for mutational analysis in urine, a bodily fluid that is readily available and abundant. However, quantification of peripheral fluid mutDNA may prove difficult to interpret, with additional factors such as the hydration state, renal disease and presence of bladder outflow obstruction in the case of urine, potentially affecting mutDNA levels. The fact that urine is freely available in large volumes, would make it of particular interest for analysis using a broader targeted sequencing approach. In addition, urinary DNA is likely to be of importance in urological cancers, especially bladder cancer, where the tumours may shed large quantities of DNA directly into the urine.

### 1.1.4 Urinary Processing

There is no consensus on urine processing for optimal urinary cfDNA yield. It is likely that the first morning urine sample will have poor DNA yield due to the prolonged exposure to urinary DNAses. Therefore, all urine samples were provided after the first morning micturition. However, 3 urine processing factors were investigated further. Firstly the influence of time between processing samples and urinary cfDNA yield was investigated. In plasma, cfDNA yields have been shown to increase when processed after from 4-25 hours after collection (Xue *et al.*, 2009). This is likely due to cell lysis, most probably leucocytes, releasing wild-type DNA into the sample. Therefore delayed plasma processing would obscure mutDNA analysis where detection of rare mutant molecules is already challenging. The effect of delayed processing on urinary cfDNA yield has not yet been investigated.

Secondly, the addition of Ethylenediaminetetraacetic acid (EDTA) has also been shown to improve the cfDNA yield from plasma for ctDNA analysis (Lam *et al.*, 2004), due to EDTA

mediated ion chelation and hence DNAse activity inhibition (Barra *et al.*, 2015). For urine samples the effect of EDTA on cfDNA remains to be investigated however, previous studies have added EDTA in concentrations ranging from 10mmol/L (Su *et al.*, 2004; Tsui *et al.*, 2012) to 25mM final concentration (Millholland, 2012).

Thirdly, the effect of centrifugation on urine was chosen for investigation. For plasma, centrifugation of samples have been suggested to improve the quality of cfDNA for rare mutant molecule analysis (Lam *et al.*, 2004). In urine, there are likely to be less leucocyte to contaminate samples. However, lysis of shed urothelial cells could contaminate mutant molecule analysis, though may improve mutant molecule analysis in bladder cancer. Previous urinary cfDNA reports have often used a low speed, single centrifugation step for 10mins (Su *et al.*, 2004; Tsui *et al.*, 2012).

### 1.1.5 Size of Urinary cfDNA

Plasma cfDNA has a 166bp peak size with further peaks periodically every 10bp (Lo *et al.*, 2010), thought to correspond with a turn of the DNA helix around a histone core (Mouliere and Rosenfeld, 2015). At the start of my investigations, there were few reports investigating urinary cfDNA size. Of note, Su *et al.* performed size analysis on urinary DNA using polyacrylamide gels and found two main bands of DNA (Su *et al.*, 2004). The first a band approximately 150 – 250 nucleotides long, which they hypothesised derived from the circulation through renal filtration, the second a much larger band representing DNA >1kilobase long, which they hypothesised derived from the lining of the urinary tract (Su *et al.*, 2004). In a subsequent study by Tsui *et al.*, greater resolution NGS approaches were used to assess the size of DNA present in the urine of pregnant women. They found that maternal urinary cfDNA was had a predominant peak at 29bp and the majority of maternal cfDNA was below 100bp in length (Tsui *et al.*, 2012). If a similar difference in mutDNA and cfDNA size distribution is present in urine for cancer patients, size selection (either during library preparation or in-silico) may help select mutDNA and aid their investigation.

## 1.2 Prostate Cancer

Prostate cancer has affected men since antiquity. Though Adams demonstrated prostate cancer pathology for the first time in 1853, the clinical entity of prostate cancer has been described since the time of the ancient Egyptians. Indeed, the first known human prostate cancer was detected in “M1”, an Egyptian mummy, using High Resolution Computerised

Tomography (Prates *et al.*, 2011). In the 1950's, men with prostate cancer would be typically diagnosed when they presented with bony metastasis (Denmeade and Isaacs, 2002). Since then many changes, including the advent of Prostate Specific Antigen testing has allowed for earlier diagnosis. Despite the trend for earlier diagnosis, prostate cancer remains a significant cause of mortality and improvements are required in particular in our ability to distinguish between indolent and aggressive forms of prostate cancer.

### 1.2.1 Prostate Cancer Epidemiology

Prostate cancer is the most common male malignancy. Prostate cancer is almost always Prostatic adenocarcinoma (McAninch *et al.*, 2013). Since the work of Franks in the 1950s, post-mortem studies of men have been the main method of inferring the true prevalence of prostate cancer amongst men. Of American Caucasian men who died from trauma, Sakr *et al.* demonstrated the prevalence of prostate cancer and its precursor (Prostatic Intra-epithelial Neoplasia – PIN) to be as high as 44% for men between 51 and 60 years of age, rising up to 83% in men aged between 71 and 80 year old men (Sakr *et al.*, 1993). In 2011 there were 41,736 new diagnoses of prostate cancer (CRUK, 2014). The incidence increases sharply with age. The incidence per 100,000 men is 800 for those aged between 75 to 79 compared with only 60 for men aged between 50 to 54 (CRUK, 2014). Furthermore, the incidence of prostate cancer is increasing (Hsing *et al.*, 2000; CRUK, 2014).

Prostate cancer is now the second leading cause of cancer mortality in men and caused 10,800 deaths in the U.K. in 2011 with the majority of deaths occurring in men aged over 75 years (2014). 5 year survival rates have increased from 31% during 1971 - 1975 to 68.3% during 1996 – 2000 (CRUK, 2014). For 2005 – 2009, the 5 year survival rate was 81.4% (CRUK, 2014) and there were an estimated 10.07 deaths per 100,000 men in the U.K. (Haas *et al.*, 2008).

Prostate cancer is staged according the TNM classification system. The major "T" or Tumour stages are:

- T1 tumour is too small to be felt during digital rectal examination
- T2 tumour is confined to the prostate gland
- T3 tumour invades prostatic capsule
- T4 tumour invades surrounding organs

Historically, most men presented with late stage complications of prostate cancer and life expectancy from time at diagnosis was short. The recent improvement in survival rates has largely been due to improvement in early diagnosis of prostate cancer and treatment.

### 1.2.2 Prostate Cancer Diagnosis

The current trend for early diagnosis in prostate cancer is largely due to the increase utilisation of Prostate Specific Antigen (PSA) testing. PSA is a kallikrein-related serine protease normally found in prostatic secretions but found in blood when the normal prostatic architecture is disturbed (Lilja *et al.*, 2008). Since its use as a screening tool from the 1980's PSA-detected early prostate cancer has become more prevalent. However, recent studies suggest that early treatment of all prostate cancers may not be necessary. In one study of >180,000 men with prostate cancer, Schroder *et al.* found that 1410 men would need to be screened and 48 additional cases of prostate cancer would need to be treated to prevent one prostate cancer death (Schroder *et al.*, 2009). Indeed, in the PIVOT trail, Wilt *et al.* report no significant difference between men treated with radical prostatectomy and those who underwent observation in 731 men. However, when categorised according to risk, men with high-risk prostate cancer had improved all-cause mortality following prostatectomy (Wilt *et al.*, 2012). Therefore, the impact that PSA testing and subsequent treatment of localised disease has had on prostate cancer specific mortality (PCSM) is questionable (Schroder and Zappa, 2012; Andriole *et al.*, 2012) and there is therefore a drive to manage men with less invasive therapies or active surveillance (a PSA & biopsy based protocol) alone. Indeed, several international guidelines suggest the use of active surveillance for prostate cancer at low risk of causing PCSM (Heidenreich *et al.*, 2014; NICE).

#### 1.2.2.1 Biopsy approach

The diagnosis of localised prostate cancer relies on the histological analysis of prostate biopsy tissue samples. The original random sextant systematic prostate biopsy involved a 6-core biopsy taken trans-rectally with the help of ultrasound guidance (Hodge *et al.*). In 2004, King *et al.* demonstrated that a 10-core transrectal biopsy, with an additional 2 lateral-mid and 2 lateral base biopsies had final pathological upgrading rates of 13% compared with 25% of men undergoing sextant biopsy when compared to surgical pathology in 78 men. Interestingly however, there was no statistical difference between the rates of men who had correctly identified grading status (62% for sextant and 63% for 10-

core biopsy) (King *et al.*). More recently however, there has been the inclusion of ever more cores (Moore *et al.*) and saturation (>20 cores) rectal or transperineal biopsies have been advocated to increase the rate of clinically significant prostate cancer detection (Guichard *et al.;* de la Taille *et al.*) and by proxy therefore to reduce the upgrading rate from final pathology. However, in a systematic review, Eichler *et al.* demonstrated that the addition of further cores (18 to 24) did not detect significantly more cancers and may have had a poorer side-effect profile, though reporting for reviewed studies was poor (Eichler *et al.*). Currently therefore, 10-12 core TRUS biopsy is standard practice in the initial biopsy setting, despite having been shown to miss around 30-45% of prostate cancers (Scattoni *et al.*, 2007; Lecornet *et al.*, 2012). Indeed, a recent survey of Urologists in the U.K. revealed that saturation biopsy is infrequently used in the initial biopsy setting, only when transrectal biopsies are contraindicated or if the prostate volume is >70cc (NICE, 2014).

Biopsy techniques continue to evolve and transperineal template biopsy strategies, coupled with multi-parametric MRI guidance (MRI-B) are increasingly being adopted. Pinto *et al.* used fusion MRI-B to demonstrate an improved efficiency of diagnosis (fewer number of cores needed to detect prostate cancer). However, they were unable to demonstrate a change in the rate of detection (Pinto *et al.*). In 2013, Moore *et al.* conducted a systematic review of MRI-B and standard TRUS biopsy and similarly revealed no difference in the detection of clinically significant cancer whilst confirming the ability of MRI-B to detect cancer with fewer cores (Moore *et al.*, 2013). These initial results are exciting however, studies determining the value of MRI-B compared with TRUS biopsy in relation to long-term outcomes are needed.

#### 1.2.2.2 Histological classification

Since 1966, the Gleason Grading has summed the two most common grade patterns, each scored from 1-5 according to glandular architecture, to produce a Gleason score. In 2005 the International Society of Urological Pathology (ISUP) updated the Gleason scoring system (Epstein *et al.*, 2005). The 2005 ISUP changes were broadly aimed at limiting the scope of glandular architecture pattern 3 whilst widening the scope of pattern 4 (Montironi *et al.*, 2010; Epstein *et al.*, 2005). Therefore cancers that were previously graded 3 were subsequently graded as 4. Indeed, Greenburg *et al.* demonstrated that adoption of 2005 ISUP scoring led to an increased incidence of intermediate and high-risk prostate cancer whilst low-risk prostate cancer incidence remained stable in the U.K. In 2014, the ISUP committee updated guidelines to grade prostate cancer, largely to overcome the fear that

patients had when assigned a cancer diagnosis but with a Gleason score of 6 (Epstein *et al.*, 2016a). The novel Grade Groups assigns a prognostic score of 1-5 according to glandular pattern (see Table 1). The lead author showed these groups to be a more accurate predictor of BR in American men undergoing radical treatment (Epstein *et al.*, 2016b). Most recently, a new risk stratification system incorporating this new prognostic grouping system has been shown to significantly improve prediction of PCSM in a new diagnosis cohort of U.K. men (Gnanapragasam *et al.*, 2016).

### 1.2.3 Risk stratification following prostate cancer biopsy

Risk stratification tools have repeatedly been shown to outperform clinicians at identifying the men at low risk of PCSM and in whom more conservative management strategies should be pursued (Shariat *et al.*, 2009). Men with localised prostate cancer are currently risk-stratified according to their PSA at diagnosis, clinical stage and Gleason grade at biopsy. In the U.K., the National Institute for Clinical Excellence (NICE) has published a scoring system that groups men according to the risk of prostate cancer recurrence following treatment. In brief, there are three groups: low-risk, for men with a PSA <10ng/ml, Gleason score ≤6, and clinical stage T1-T2a, the intermediate-risk group for men with a PSA 10–20ng/ml, or Gleason score 7, or clinical stage T2b and a high-risk group for men with a PSA >20ng/ml, or Gleason score 8-10, or clinical stage ≥T2c. However, these guidelines draw on studies that have never been assessed against PCSM. Indeed, the PSA cut-off points are replicated from D'Amico's work in 1998, using biochemical recurrence (BR) as a surrogate for aggression which, does not always predict PCSM (Freedland *et al.*, 2005; Pound *et al.*, 1999). Another early attempt at allocating risk, the Kattan nomograms predicted the presence of indolent cancers based on a study of clinical stage, Gleason grade, PSA and cancer volume in biopsy specimens (Kattan *et al.*, 1998). However, the majority of patients in this study underwent radical prostatectomy and again outcomes were measured against BR. Furthermore, clinical staging has not consistently been shown to associate with PCSM in the U.K. For example, although Reese *et al.* demonstrated that clinical staging predicted BR after radical prostatectomy in American men (Reese *et al.*, 2012), Selvadurai *et al.* showed that clinical staging could not predict adverse histology on repeat biopsy for U.K. men undergoing active surveillance (Selvadurai *et al.*, 2013). There have been no U.K. based studies to show that the above criteria can predict PCSM.

Despite this, with some relatively minor changes, the above clinical parameters are incorporated in over 20 risk stratification tools (Rodrigues *et al.*, 2012). Many of these tools are validated in populations of men outside the U.K., who are undergoing radical prostatectomy and are measured against surrogate markers other than PCSM.

In 2005 Cooperberg *et al.* described the Cancer of the Prostate Risk Assessment (CAPRA) score. The score indicates risk on a 1-10 scale by using; age and PSA at diagnosis, percentage of biopsy cores which contain cancer, Gleason score at biopsy and clinical stage, to assign points. Although the score was initially validated against predicting BR following radical prostatectomy, it has subsequently been shown to predict risk of bone metastases and PCSM in over 10,000 American men with localised cancer considering all treatment options. Cooperberg *et al.* showed that for each increase in CAPRA score, there was a statistically significant increased risk of prostate cancer specific morbidity or mortality (Cooperberg *et al.*, 2009). However, only 6% of men undertook active surveillance / watchful waiting in their cohort, despite 49% of patients having a CAPRA score of 2 or less.

National and International consortia have also acted on evidence to suggest that increasing the number of risk categories can improve pre-treatment risk stratification. The National Comprehensive Cancer Network (NCCN) have recently updated their risk stratification guidelines to incorporate the amount of disease present in prostate biopsy cores and include a new very-low risk group. The purpose of the very low risk group being to highlight patients who are suitable for surveillance strategies rather than aggressive treatment. The drive for creation of the new group came from studies such as that conducted by Miller *et al.*, showing that approximately 40% of cancers diagnosed in USA were over-treated (Miller *et al.*, 2006). Furthermore, in 2008 Beasley *et al.* showed that splitting the intermediate group into low-intermediate and high-intermediate groups also improved pre-treatment risk stratification. However, these conclusions were due to the high-intermediate group having reduced BR rates when given adjuvant androgen deprivation therapy whilst the low-intermediate group received no benefit (Beasley *et al.*, 2008; Rodrigues *et al.*, 2012).

NICE therefore, highlights that clinical risk stratification tools may not be representative of outcomes in U.K., not least because of the differences between PSA screened populations and ours, and differences in the way men are treated (NICE, 2014). A recent study from our institution has addressed this and tested the NICE risk groups and a new model in a large U.K. population. This study demonstrated poor concordance for the NICE groups in

predicting PCSM but a significantly improved performance by using a more refined risk stratification system (Gnanapragasam *et al.*, 2016).

But, there are multiple methods to stratify risk in prostate cancer, including nomograms, UCSF-CAPRA Score and D'amico classification. In brief these scoring methods take several factors such as Gleason score, PSA level, and clinical stage to calculate risk. Of these the D'amico classification, first described in 1988, is the most commonly used in practice (D'Amico *et al.*, 1998). Despite extensive use, many men with seemingly low risk prostate cancer can still die of their disease. In a series of 458 men with low risk prostate cancer being treated with active surveillance, Klotz *et al.* found that 30% of men had a PSA doubling time of <3yrs (another contentious marker of high risk cancer) or had upgrading on repeat biopsy and were therefore offered definitive treatment (Klotz *et al.*, 2010).

Concurrently, there is a real risk of being under staged during standard TransRectal Ultrasound Guided prostate biopsy. During analysis of 499 radical prostatectomy patients, Steinberg *et al.* found that only 58% of their academic hospital patients and 34% of patients from community hospitals had the same Gleason score when compared to their initial needle biopsy grade (Steinberg *et al.*, 1997). However, a more recent study suggests that concordance is now around 75%, likely due to the increased number of cores taken (Rajinikanth *et al.*, 2008). The non-invasive analysis of prostate cancer genomes through mutDNA analysis may aid in accurately differentiating between indolent and aggressive prostate cancer if it represents mutations from multiple (or the most significant) clone(s).

### 1.2.4 Prostate Cancer Mutational Profile

WGS and WES studies have demonstrated that prostate cancer has few mutations. Indeed, the mutation rate in localised disease is <0.9 mutations/Mb (Berger *et al.*, 2011) and 2 mutations /Mb for CRPC (castrate resistant prostate cancer) (Grasso *et al.*, 2012). By contrast, the mutation rate of other solid cancer such as malignant melanoma can be as high as 15 mutations /Mb (Hodis *et al.*, 2012).

Of the mutations that are present in localised disease, few are recurrent except the TMPRSS:ETS re-arrangement, which is present in >50% of all patients (Tomlins *et al.*, 2005; Berger *et al.*, 2011). Unfortunately the intergenic location of the breakpoint for this re-arrangement is not recurrent and therefore non-invasive biomarkers targeting this

recurrent fusion in circulating nucleic acids have been limited to mRNA testing. Though fused mRNA fragments can be detected in the urine of men with prostate cancer (Laxman *et al.*, 2006) and initial lab based tests were encouraging, sensitivity in the clinic was low (Hessels *et al.*, 2007), possibly due to the fact that only 50% have this re-arrangement and due to the inherent instability of single stranded nucleic acids. For SNV's, the most common mutations are *TP53*, *SPOP, PTEN* and *FOXA1* which are only found in 14%, 8%, 7% and 5% of men respectively (Forbes *et al.*). Furthermore, the prevalence of these mutations are significantly affected by the disease setting, with clones harbouring *SPOP* SNV's often lacking TMPRSS:ETS re-arrangements (Barbieri *et al.*, 2012). The low mutation rate will make unbiased amplicon sequencing methods to detect ctDNA challenging and may require broader sequencing strategies.

### 1.2.5 Prostate heterogeneity obfuscates accurate stratification in localised disease

Vogelstein *et al.* summarised the mutational profile of multiple tumour types to define four categories of tumour heterogeneity (Vogelstein *et al.*, 2013):

1. Intra-tumoural: heterogeneity among the cells of one tumour.
2. Inter-metastatic: heterogeneity among different metastatic lesions in the same patient.
3. Intra-metastatic: heterogeneity among the cells of an individual metastasis.
4. Inter-patient: heterogeneity among the tumors of different patients.

Prostate cancers represent an extremely heterogeneous tumour in all four regards. Indeed, Haffner *et al.* performed tumour re-sequencing on multiple tumour clones in a single prostate cancer patient to find 9 distinct lesions. Of notable interest, by performing tumour re-sequencing of metastatic sites following rapid autopsy, Haffner *et al.* demonstrated that the metastatic clone hailed from a small tumour with Gleason 3+3 grading as opposed to the surrounding higher Gleason tumour clones (Haffner *et al.*, 2013). More recently, Cooper *et al.* have corroborated that localised PC can be multi-focal by using multi-regional WGS to assess regions 6-8 regions of prostatectomy specimens from 3 men (Cooper *et al.*, 2015). Of interest, they found that many cancer regions contained distinct clones each with independent genetic phylogenies.

Improvements in risk stratification are urgently needed and are likely to be borne of the heterogeneous nature of localised prostate cancers. These factors translate to real uncertainty in clinical practice and current methods to overcome heterogeneity include

increasing the numbers of biopsy cores being taken (Boyd *et al.*, 2012). There is therefore a need for a biomarker than can monitor tumour heterogeneity and progression non-invasively.

### 1.2.6 Biomarkers using genetic analysis

There has been an increase in the use of genetic analysis to predict outcomes in prostate cancer. However, the majority of these studies have used prostatectomy samples, where the amount of tissue available makes genetic analysis easier. Through advances in PCR methods and genetic analysis, there is now the real ability to detect variants from the small amounts of genetic material found in bodily fluid or biopsy specimens. For biopsy specimens, several diagnostic aids are available e.g. Oncotype Dx (Knezevic *et al.*, 2013), Prolaris (Cuzick *et al.*, 2012), Dechipher (Klein *et al.*), etc. However, many of these assays assess the likelihood of aggressive cancer by analysing the expression of genes from isolated RNA, which can be unstable in bodily fluids.

The presence of ctDNA has previously been reported by Bettegowda *et al.*, who detected ctDNA in 2/5 patients with metastatic prostate cancer (Bettegowda *et al.*, 2014) and by Heitzer *et al.*, who described Plasma-Seq, a sWGS (shallow whole genome sequencing) approach to detect CNAs (copy number aberrations) in plasma of patients with CRPC (Heitzer *et al.*, 2013). Heitzer *et al.* performed Plasma-Seq on 13 plasma samples to detect CNAs in 80% of patients when the AF is >10%. More recently, Carreira *et al.* demonstrated the use of deep amplicon sequencing and WGS to track SNVs and CNAs in sequential plasma samples from 16 ERG positive CRPC patients (Carreira *et al.*, 2014). Despite this, there are no reports of ctDNA analysis in localised PC, where ctDNA levels are likely to be much less, potentially rendering WGS type approaches less useful. Furthermore, if the ability of ctDNA to monitor clonal evolution in metastatic disease can be applied to localised PC, a ctDNA based test could lead to profound clinical benefit.

## 1.3 Bladder Cancer

The earliest descriptions of bladder cancer can also be found from scriptures dating back to ancient times. The first documented European report of bladder cancer was not until an oral account in 1895 by Ludwig Rehn (Dietrich and Dietrich, 2001). This ignited research into the aetiology of bladder cancer. Historically, this remained the focus of research efforts until recently. Despite the breadth of research currently investigating bladder

cancer, it accounts for 3% of all cancer deaths (CRUK, 2014) and is the most expensive solid cancer to treat per patient, due to its high recurrence rate leading to costly in and outpatient hospital admissions (Botteman *et al.*, 2003; Yeung *et al.*, 2014).

### 1.3.1 Bladder Cancer Epidemiology

Bladder Cancer is the most common malignancy of the urinary tract, and the 7$^{th}$ most common cancer overall in the U.K. The risk of developing bladder cancer is strongly affected by occupational hazards, infection and environmental factors. For example in Egypt, the age-standardised incidence of bladder cancer is 21.8 per 100,000 men and 5.6 per 100,000 women, whilst in the U.K, the age-standardised incidence of bladder cancer is 9.3 per 100,000 men and 2.8 per 100,000 women (Ferlay *et al.*). There were 10,339 new cases diagnosed in 2011 in the U.K. and incidence rates have been decreasing overall since the early 1990s (CRUK, 2014). The incidence in men is higher than in women by a ratio of 2.5 male : 1 female in 2011 (ONS, 2011).

Primary bladder cancer has several pathological subtypes (see Table 1.1) and the incidence of these subtypes vary depending on geographical location. Historically, the prevalence of *Schistosomiasis* infections in Egypt was high. As chronic *Schistosoma Haematobium* infections are associated with squamous cell carcinoma (SCC) (Mostafa *et al.*, 1999), during the 1980's the dominant subtype of bladder cancer was SCC accounting for 78% of the disease (Fedewa *et al.*, 2009). More recently, following public health initiatives and changes to the habitat of *Schistosoma Haematobium,* there has been a reduction in the prevalence of Schistosomiasis in Egypt making urothelial carcinoma more prevalent (Fedewa *et al.*, 2009). In the U.K., smoking, and industrial exposure are the main risk factors and therefore Urothelial carcinoma is the dominant pathological subtype accounting for 90% of disease (Fedewa *et al.*, 2009). For the purposes of my research I will therefore focus on urothelial carcinoma of the bladder and where bladder cancer is stated, it will describe urothelial carcinoma of the bladder.

| Malignant | Frequency in UK, % |
|---|---|
| Urothelial carcinoma (previous Transitional Cell Carcinoma) | 94.72 |
| Urothelial carcinoma variants | |
| • Keratinising | 0.01 |
| • Squamous differentiation | 0.25 |
| • Nested | <0.01 |
| • Carcinosarcoma | 0.16 |
| • Sarcomatoid | 0.79 |
| • Undifferentiated carcinoma | 0.03 |
| • Signet ring | 0.02 |
| • Clear cell | 0.01 |
| • Inverted papilloma | 0.04 |
| Squamous cell carcinoma (SCC) | 1.65 |
| Adenocarcinoma | 1.61 |
| Mesenchymal | 0.91 |
| Haematopoetic / lymphoid | 0.04 |
| Neuroendocrine | 0.24 |
| **Secondary Malignancies** | **0.08** |

**Table 1.1: Bladder cancer histological subtypes, modified from (Sexton *et al.*, 2010) and (Boustead *et al.*, 2014)**

During 2005 - 2009, 56.1% of adult patients with bladder cancer in England survived their cancer for five years or more (ONS, 2011). The survival rate has decreased recently; the five year survival for English men during 1991-1995 was 64.2% and is now down to 58.2% (CRUK, 2014). This is partly to be due to the recoding of carcinoma in-situ and other lesions that were historically coded as cancerous (Shah *et al.*, 2008).

### 1.3.2 Aggressive & Non-Muscle Invasive Bladder Cancer

Bladder cancer is staged according to the TNM classification system. The major “T” or Tumour stages are:

- Tis / Cis tumour cells remain only in the epidermal layer and have not breached the basement membrane
- T1 tumour cells are found in the subepithelial connective tissue
- T2 tumour invades muscularis propria
- T3 tumour invades perivesical adipose tissue
- T4 tumour invades surrounding organs

Tis - T1 can be termed non-muscle invasive bladder cancers (NIMBC) and represent of 80% of urothelial carcinoma at first diagnosis. The remaining 20% are muscle invasive bladder cancers (MIBC) and comprise of T2 - T4 (Kirkali *et al.*, 2005).

The World Health Organisation classifies bladder cancer into the following grade groupings (Kirkali *et al.*, 2005):

- Papilloma Benign tumour
- PUNLMP Papillary urothelial neoplasm of low malignant potential
- Low grade Cells retain an orderly appearance with minimal disruption to structural architecture
- High grade Highly abnormal & disordered cells

Most NMIBC is low grade at diagnosis, with only 4% being high grade (Sylvester *et al.*, 2005). Even after complete excision the recurrence risk is high for both low and high grade NIMBC, with low grade NMIBC recurring in 64% and high grade NMIBC recurring in approximately 56% of patients (Kirkali *et al.*, 2005). NMIBC survival rates depend on the risk of the urothelial cancer invading the muscle layer and becoming invasive bladder cancer. Once urothelial cancer invades the muscle layer of the bladder, the risk of metastasis increases significantly and, metastatic bladder cancer has a poor 10% survival rate over 5 years. Therefore, the risk of NMIBC progressing drives clinical decision-making.

Several factors determine the risk of progression, including tumour grade. The risk of low grade NMIBC progressing is 10.5% whilst for high grade NMIBC the risk is 27.1% (Kirkali *et al.*, 2005). Sylvester *et al.* used data from previous European Organization for Research and Treatment of Cancer (EORTC) studies retrospectively determine a scoring system based an analysis of 2596 patients. The system allocated scores to the features of the tumour including; number of tumours, tumour size, prior recurrence rate, stage, presence of CIS and grade to determine the risk of recurrence and progression. Using the EORTC

scoring, the most weighted scores for determining progression are allocated to high-grade tumours and those with CIS present.

High grade and CIS tumour share a common genetic profile, with mutations in *TP53* and *RB* a common feature. These tumours often go on to progress to life threatening metastasis (Wu, 2005). Conversely, low-grade papillary tumours are often slow to metastasis. These tumours are more likely to have mutations in the receptor tyrosine kinase–Ras pathway, particularly in DNA coding for the Fibroblast growth factor receptor protein FGFR3 or in *HRAS* (Wu, 2005). Both tumours harbour deletions in Chromosome 9 from an early stage.

### 1.3.3 Treatment Options For Muscle Invasive Bladder Cancer

As the progression risk is high despite aggressive NMIBC being localized, doctors and patients face a difficult decision regarding what treatment to take to ameliorate the risk of progression. There are broadly speaking two commonly used treatment options.

#### 1.3.3.1 Radical Radiotherapy

Ionising radiation has been used to treat cancer since the late 19$^{th}$ Centaury, initially with Radon, then cobalt x-rays. Modern radiotherapy is delivered by linear accelerators, which focus a beam of ionising radiotherapy to the tumour and surrounding structures. For MIBC, it is most often delivered in fractions of ~2Gy /day, 5 days a week to a total of 45-50Gy. More recently, improvements in radiotherapy planning, verification, and delivery have allow more targeted delivery of higher radiation doses, with less side effects (Zhang *et al.*, 2015). Using hypofractionated intensity-modulated radiation therapy, Turgeon *et al.* report a 71% 3 year survival (Turgeon *et al.*, 2014).

#### 1.3.3.2 Radical Cystectomy

Cystectomy has been performed since the 1800s. The initial procedure was more akin to what would be termed a simple cystectomy today, with no attempt to control perivesicular organs or regional lymph nodes (Raj and Bochner, 2007). Despite this, the procedure had a formidable mortality rate in excess of 30% (Raj and Bochner, 2007). In 1949, following their investigation into local recurrence by cadaveric dissection, Marshall and Whitmore described the prototypical radical cystectomy – surgical removal of the bladder, adjacent organs and their lymphatic drainage. Through advances in surgical technique, anaesthesia

and perioperative care, modern radical cystectomy has a 30 day mortality rate of 2% and following radical cystectomy for organ confined cancer, a disease specific five year survival of 94% (Manoharan *et al.*, 2009).

### 1.3.4 Neoadjuvant Chemotherapy for Muscle Invasive Bladder Cancer

Bladder Cancer (BC) is the most common malignancy of the urinary tract (Ferlay *et al.*, 2013). Approximately 25% of BC patients have Muscle Invasive Bladder Cancer (MIBC) at diagnosis (Kirkali *et al.*). The presence of MIBC signifies aggressive disease with a significant risk of metastatic progression. Even after definitive treatment (e.g. radical cystectomy) overall survival (OS) is on average 50% over 5yrs (Stein and Skinner). Meta-analyses show that a 3-month regimen of cisplatin-based neo-adjuvant chemotherapy (NAC) can improve 5-year absolute OS by 6% (Vale, 2003; Trialists, 1999; International Collaboration of *et al.*, 2011).

Despite maximum treatment with NAC and radical cystectomy, many patients do not respond to NAC (Lee *et al.*) and MIBC patients often recur within 2 years (Koie *et al.*, 2015). Currently biomarkers to predict outcome during NAC are lacking and the predictive capacity of cystoscopy or radiological examination is limited (Nishimura *et al.*). In patients not responding to NAC, definitive treatment could be expedited or other systemic treatments investigated in the neo-adjuvant setting. Of note, the recent success of immune checkpoint inhibitors (Rosenberg *et al.*) provides a possible alternative non-cross-resistant systemic treatment for these patients and trials with checkpoint inhibitors have recently been initiated in the peri-operative setting.

Circulating tumour DNA (ctDNA) offers a minimally-invasive means to monitor tumour status. The half-life of circulating cell-free DNA is reportedly less than 2 hours (Diehl *et al.;* Fournié *et al.;* Thierry *et al.*, 2010) and the allele fraction ratio of mutant:wild type DNA (AF) has been shown to reflect tumour burden and response (Rago *et al.*, 2007; Thierry *et al.*, 2010; Dawson *et al.*). The translational potential for mutDNA in body fluids could be even greater in bladder cancer due to the possibility of monitoring mutDNA in urine (Patel and Tsui, 2015), a peripheral fluid that truly can be collected ‘non-invasively’. In BC, the close proximity of tumour to the peripheral fluid reservoir might be expected to lead to a greater accumulation of tumour derived DNA.

Recent reports have shown that mutDNA is detectable in the plasma and the urine of patients with BC. Using the Affymetrix Oncoscan assay, Togneri *et al.* demonstrated that

mutation profiles present in FFPE tumour specimens were mirrored in matched urinary samples from patients with Non-Muscle Invasive Bladder Cancer (Togneri *et al.*). Ward *et al.* used digital PCR (dPCR) and NGS to analyse the somatic mutation status of cell pellets obtained by centrifuging urine samples (UCP), and urine supernatant (USN) in 120 primary patients with bladder cancer and 89 patients post transurethral resection (TUR), and detected mutDNA in 70% of cases (Ward *et al.*). Furthermore, Birkenkamp-Demtröder *et al.* detected mutDNA in plasma and urine of patients with BC by sequencing tumour specimens to design personalised droplet dPCR probes for use in peripheral fluids. In 4/6 patients, personalised probes detected plasma ctDNA several months before clinical progression (Birkenkamp-Demtroder *et al.*). These studies mostly focus on non-invasive BC, which are often driven by different pathways than MIBC (Cancer Genome Atlas Research, 2014), and none compared plasma, UCP and USN sampling methods. The application of mutDNA in BC represents an exciting opportunity for clinical impact in MIBC and, despite recent efforts, remains relatively unexplored.

# CHAPTER 2: URINARY CELL-FREE DNA

## 2.1 Synopsis

Analysis of mutant cfDNA has the potential to help predict aggressive disease, monitor treatment response and monitor tumour burden. For urological cancers, the added potential is that mutated cfDNA may be locally shed into a body fluid that is both abundant and truly non-invasive. Indeed, Millholland *et al.* demonstrated the presence of *FGFR3* hotspot mutations in urinary cfDNA, though their approach was limited by single gene analysis (Millholland, 2012). At the initiation of this project, there were no direct comparisons of urinary cfDNA processing or extraction methods and no multi-gene sequencing based reports of urinary mutDNA analysis.

To help unlock the potential of urinary cfDNA as a liquid biopsy, 7 healthy volunteer urine samples were used to investigate urine cfDNA extraction and processing. Four urine specific DNA extraction kits were compared for cfDNA yield. Urine samples were processed with or without centrifugation, and the addition of EDTA. To investigate the affect of increased length of time between harvesting and processing, urine samples were processed at 1hour, 6hours and 48hours after collection. A further 2 patients with metastatic bladder cancer were recruited to assess optimal mutDNA approaches for BC using a novel bladder specific primer panel.

Of the four kits that we assessed, mean urinary cfDNA across the samples was 22,810, 14,570, 7,243 and 1,535 Genomic Equivalent (GE) copies/ml sample for QIAmp circulating nucleic acid, Norgen, NeoGenStar and SNOVA kits respectively). Only the SNOVA urinary cfDNA extraction kits had a statistically significantly lower cfDNA yield ($p<0.01$). We found that centrifugation reduced the urinary cfDNA yield ($p<0.001$) and hypothesise that this is through removal of the fraction of urinary cfDNA that would arise from cellular lysis during extraction. The addition of EDTA did not significantly alter the cfDNA yield, however there was a small effect to improve cfDNA yield with the addition (19,350 versus 15,170 mean GE copies/ml for samples with 10mMol EDTA compared to no EDTA). Increased length of time between sample collection and processing reduced the cfDNA yield ($p=0.0121$). Using the bladder-specific panel, no SNVs were detected from the metastatic BC patients, precluding further analysis of urinary processing for mutDNA.

These results demonstrate the feasibility of urine processing for cfDNA analysis using a scalable approach. Urinary cfDNA has since been shown to be an important source of

mutDNA in BC (Millholland, 2012; Togneri *et al.*, 2016), colorectal cancer (Su *et al.*, 2004) and others. The optimisation of urinary cfDNA extraction will facilitate the application of mutDNA analysis in the Rosenfeld laboratory.

## 2.2 Publications arising from this work

*Work presented in this chapter was first described in my first year report and parts were published in Clinical Biochemistry as well as sections of the supplementary material of a published in Scientific Reports. The text is therefore excerpted from my first year report and related publications except to display updated data and references.*

**K.M. Patel**, D.W.Y. Tsui. The translational potential of circulating tumour DNA in oncology. Clinical Biochemistry. October 2015. 48:15, 957-961. ISSN 0009-9120, http://dx.doi.org/10.1016/j.clinbiochem.2015.04.005.

**KM Patel**[#], KE van der Vos[#], CG Smith[#], FC Mouliere, D Tsui, J Morris, D Chandrananda, F Marass, D van den Broek, DE Neal, VJ Gnanapragasam, T Forshew, BW van Rhijn, CE Massie, N Rosenfeld*, MS van der Heijden*. Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. Scientific Reports. 2017. 7:5554. http://dx.doi.org/10.1038/s41598-017-05623-3

[#]These authors contributed equally.

* These authors are joint senior supervisors

## 2.3 Aims

My primary objective was **to explore whether urinary mutant cfDNA can be isolated for future investigation as an informative marker for prostate & bladder cancers**. To develop methods for urinary cfDNA analysis, I:

1. **Initiated sample collections:** I collected fresh urine from healthy volunteers (men with raised PSA who have normal biopsy results attending the prostate diagnostic clinic) and from patients with metastatic bladder cancer who attended oncological clinics, using previously granted ethical approval (MREC:01/4/061, LREC:11/H0311/2 and LREC:03/018).
2. **Optimised protocols for DNA extraction from urine supernatant:** I performed literature searches to identify then test commercially available urinary cfDNA extraction kits. I used dPCR to assess the genomic yields of each kit on urine samples.
3. **Optimised protocols for processing urinary supernatant from fresh urine:** I collected urinary specimens with multiple processing techniques and quantified extracted DNA with dPCR.
4. **Test urinary mutDNA detection in pilot metastatic bladder cancer cohort:** Using the COSMIC database (Forbes *et al.*, 2011) and results of major sequencing projects (Cancer Genome Atlas Research, 2014), I compiled a list of commonly mutated genes in bladder cancer. I validated the primers using control genomic DNA and applied it to perform TAm-Seq on urinary cfDNA from metastatic patients to determine the optimal conditions for mutant urinary cfDNA detection.

## 2.4 Methods

### 2.4.1 Urinary processing protocols

Urine samples from 7 volunteers with raised PSA but negative prostatic biopsies were collected to optimise urinary supernatant processing and extraction procedures. Samples were collected under ethical approval previously granted for the investigation of biomarkers in urological diseases (LREC 03/018), see . Urine samples were processed according to Table 2.1 for the first 6 patients (Patient A – Patient F). Samples marked EDTA (dipotassium Ethylenediaminetetraacetic acid) had 600μL of 0.5M K2 EDTA per 30ml of urine added immediately to give a final concentration on 10mM. Where "centrifuged" is denoted, samples were centrifuged at 3,000rpm for 10 minutes and the supernatant kept for further analysis. Samples were processed at the indicated times post sample collection. Urine from the 7$^{th}$ patient (patient G) was stored in aliquots that had been processed within 1 hour and had 10mM EDTA added to the sample to provide multiple replicates for DNA extraction testing.

| | **Centrifuged (ml)** | **Uncentrifuged (ml)** |
|---|---|---|
| 1hr Freeze + EDTA | 30ml | <30ml |
| 1hr Freeze | 5ml | 5ml |
| 6hr Freeze + EDTA | 5ml | 5ml |
| 6hr Freeze | 5ml | 5ml |
| 48hr Freeze + EDTA | 5ml | 5ml |
| 48hr Freeze | 5ml | 5ml |

**Table 2.1: Urinary DNA sample processing variables tested.** Fresh urine samples were aliquoted into 3 processing groups according to time. Each urine sample was then processed using the combinations outlined above. + EDTA signifies the addition of 10mMol of dipotassium EDTA pH 8.0 (Invitrogen) to each urine sample. Centrifugation was for 10mins at 3000g.

Initial investigation of urine DNA testing was performed using 3 commercially available and hence scalable, DNA extraction kits. Kits were chosen based on the presence of a specific urinary cfDNA extraction protocol, their reported DNA (size specific) yield and the amount of urinary input required (i.e. between 1 and 10ml). Kit identification literature and internet

searches were performed in early 2013. Three kits were eventually selected for further testing; Norgen urine slurry kit, QIAmp Circulating nucleic acid kit, SNOVA nucleic acid kit.

### 2.4.2 Extracted DNA quantification through digital PCR

Digital Polymerase Chain Reaction (dPCR) dilutes the original sample and distributes it to multiple PCR reactions wells. The dilution is such that ideally there would be one molecule of DNA every two or more wells. Therefore, each well will have 1 or less molecules of DNA in the target region (Vogelstein and Kinzler, 1999), with only a very small number of wells (or no wells) with more than one target molecule. PCR reactions occur in parallel and presence of a DNA molecule can be detected by using primers and a fluorescent probe (TaqMan assay). The fluorescent probe consists of a fluorophore and a quencher in close proximity, covalently bonded to an assignable DNA sequence. During PCR, primer pairs anneal at either side of the probe and, during DNA polymerisation, Taq polymerase releases the fluorophore from the quencher due to the 5´–3´ exonuclease activity. The number of amplifiable copies in the original sample can then be determined by factoring the dilution factor, the number of positive wells, the volume of original sample used and the total volume of the sample master mix inputted. Allele-specific quantification can also be achieved using sequence-specific fluorescent probes. In order to simplify the process of setting up digital PCR, we use Biomark microfluidic digital PCR system (Fluidigm Corporation). The Biomark chips are available in a 12 and 48 panel chip, with 9,180 and 36,960 reaction chambers, respectively and both variant were used dependent on how many samples were being tested.

Urinary DNA size and concentration were analysed using 65, 97 and 360bp (base pair) assays. To reduce the risks of preferential amplification and primer dimer formation, assays were not multiplexed during dPCR. Assays were designed to target a sequence in the highly conserved *Ribonuclease P protein subunit p30* (*RPP30*) region of the human genome. Primers were designed to target the following intronic region of *RPP30*:

atactggaatttggggcacggacaggaagcgccgtggaggctcggagaaggaggaggacgtgggagtgttggaacaata
ctatctagacagagagagggtaggtgggggcagaggacatacacagcggagtggaggtggaggaggaggatagcatgg
cccaaatcgaagggaagcgaaagccttgagagacgagaacctgttttaactattggagttagtcatgggttctgtattccgtag
actgtaaagtgttggaatagagagattaaaaagaatcaagaggttcgtgtagtaacatctgctcctcagagacagttttcagctc
attgtgagttctgattactgatactgttcagaggtggtgctagaaatgacaaccacttcttggcaaagtgaatcacaccacttttta
ttgttctctttatgaagtatccgtggcgaaatgactgtccctttccactgctgccactaaatgaacgttatctctgtggtacttt

A 65bp sequence was detected by using the following Primers and TaqMan probes:

| | |
|---|---|
| RPP30_65bp_F | AGATTTGGACCTGCGAGCG |
| RPP30_65bp_R | GAGCGGCTGTCTCCACAAGT |
| RPP30_ROX | [ROX]TTCTGACCTGAAGGCTCTGCGCG[BHQ2] |

A 97bp sequence was detected by using the following Primers and Taq-man probes:

| | |
|---|---|
| RPP30_F1_97bp | TCATGGGACTTCAGCATGGC |
| RPP30_R1_97bp | GGTGAGCGGCTGTCTCCA |
| RPP30_ROX | [ROX]TTCTGACCTGAAGGCTCTGCGCG[BHQ2] |

A 360bp sequence was detected by using the following Primers and Taq-man probes:

| | |
|---|---|
| RPP30_F4_360_PM_F | CTTCAGCATGGCGGTGTT |
| RPP30_R4_360_PM_R | CAGCTGTAGGCCCCTATGCA |
| RPP30_362bp_HEX | [HEX]CCACCCAGACCATCGGGCCAC[BHQ1] |

DNA yield was calculated by introducing foreign DNA to each urine sample, immediately prior to DNA extraction. Foreign DNA was from *Xenopus Tropicalis* and is commonly used in the Rosenfeld lab for plasma DNA yield analysis (Dawson *et al.*). In brief, PCR amplification is utilised to create 170bp *Xenopus Tropicalis* DNA amplicon, with the following sequence:

GTAGGTGTCATCATGGGGAAGtccctggggcaggtggtgatcatgggatttgtagctgtttggcccatggattatc gatggtccatggttttcaggttggtttttacatgttattctctgtttccaaggttagcttagaaaaactGTGTTAGCAAAATC TATTGCCTGA

The amplicons are then quantified with dPCR using the following Taq-man probes:

| | |
|---|---|
| XenT_Spike.TaqMan_01_F | GTGATCATGGGATTTGTAGCTGTT |
| XenT_Spike.TaqMan_01_R | AAACCAACCTGAAAACCATGGA |
| XenT_Spike.Probe 6FAM | [FAM]CCCATGGATTATCG[BHQ2] |

6µL of approximately 1500 GE copies/µL *Xenopus Tropicalis* DNA was added to each sample during the first step of DNA extraction. Following the elution of DNA an estimate of

efficiency (DNA yield) is calculated by comparing the quantity of *Xenopus Tropicalis* amplicons in the eluted DNA with an extraction control sample prepared by introducing 6µL of *Xenopus Tropicalis* DNA and 44µL of elution buffer.

GE copies were calculated based on the number of estimated target wells detected by the Biomark using the following calculation:

GE copies/ml = Est. Targets x elution from extraction (µL) ÷ Sample input to mastermix (µL) ÷ vol of plasma (ml) x dead-space correction factor (6 ÷ 0.65 for 37K chip and 10 ÷ 4.6 for 12K chip x dilution factor

### 2.4.3 TAm-Seq Panel Design

To investigate the presence of urinary mutDNA, a panel was devised to target commonly mutated bladder cancer genes for application in TAm-Seq. Hotspots or entire coding regions of 8 genes based on recent WGS studies (Cancer Genome Atlas Research) and the COSMIC database (Forbes *et al.*) were selected. Genes were selected on the basis of their exon size and their mutational frequency. The final candidate genes are listed in Table 2.2. We aimed to design 48 or less primers to match the number of input assay channels on the microfluidic PCR access array chip (Fluidigm).

| Gene | Targeting | Expected mutation rate in BC | % of SNVs covered with panel | Total No of Samples | SNVs covered / total SNVs |
|---|---|---|---|---|---|
| BRAF | p.V600 | 2% | 90% | 469 | 9/10 |
| CTNNB1 | p.T41 & p.S45 | 2% | 100% | 491 | 9/9 |
| FGFR3 | p.S249, p.K652 & p.y375 | 41% | 100% | 5790 | 2384/2388 |
| HRAS | p.G12 | 7% | 99% | 1585 | 103/104 |
| KRAS | p.G12, p.G13 & p.Q61 | 4% | 100% | 1300 | 58/58 |
| NFE2L2 | p.G31 | 2% | 100% | 157 | 1/1 |
| PIK3CA | p.E545, p.G545 & p.HC1047 | 19% | 97% | 1142 | 215/221 |
| TP53 | all exons | 42% | 100% | 1335 | 555/555 |
| TERT | Promoter Hot Spots^ | 66% | 0% | 1887 | 0/1242 |

**Table 2.2: Genomic regions interrogated by TAm-Seq for SNV analysis.** Analysis of COSMIC data set was used to estimate the number of SNVs targeted by our panel and was updated on the 24th August 2016. Primers in red text (*TERT*) were not included in the final panel.

Primers previously designed by the Rosenfeld lab for ctDNA analysis, were present for *BRAF*, *HRAS*, *KRAS*, *PIK3CA* and *TP53*, with the remainder of genes requiring primers to be designed to target them. *CTNNB1* had partial coverage of a bladder cancer hotspot, so an additional primer pair was designed to improve coverage. Initially, primers were

designed using Primer 3 website (Untergasser *et al.*, 2012). Francesco Marass, a PhD student in the Rosenfeld group, developed an automated pipeline to design then tile primers, part way through the primer design stage of this project. Primer design in both techniques relied upon selecting primers based on the criteria listed in Table 2.3. The important criteria being the size as the amplicons needed to be long enough to be specific to a region in the human genome but at the same time short enough to be able to target mutDNA. In brief, the automated pipeline designed multiple such primers for a region prior to allocating each a score based on how far the criteria of each primer deviate from the median and hence calculates the optimal set of primers to tile across a region.

| Criteria | Range |
|---|---|
| primer size | 16 - 35bp |
| amplicon size | 60 - 100bp |
| annealing temperature | 55-65 |
| GC content | 45-55% |

**Table 2.3: The criteria for designing and selecting primers for targeting mutDNA**

### 2.4.3.1 *CTNNB1* primer design

*CTNNB1* is a dual function protein, regulating the coordination of cell–cell adhesion and gene transcription that is mutated in ~2% of patients with bladder carcinoma (Forbes *et al.*). The previous primer pair targeting *CTNNB1* only covered the p.S37 hotspot mutation. The p.S45 hotspot, frequently mutated in urothelial cancer, would not have been targeted, see Figure 2.1. Therefore a primer pair was designed to target the region spanning chr3:41265875-41266356 using Primer3 (Untergasser *et al.*, 2012), Figure 2.2. The region of interest is contained within square brackets and the previously designed primer targets bases, highlighted in yellow, whilst the new primer pairs are highlighted in pink.

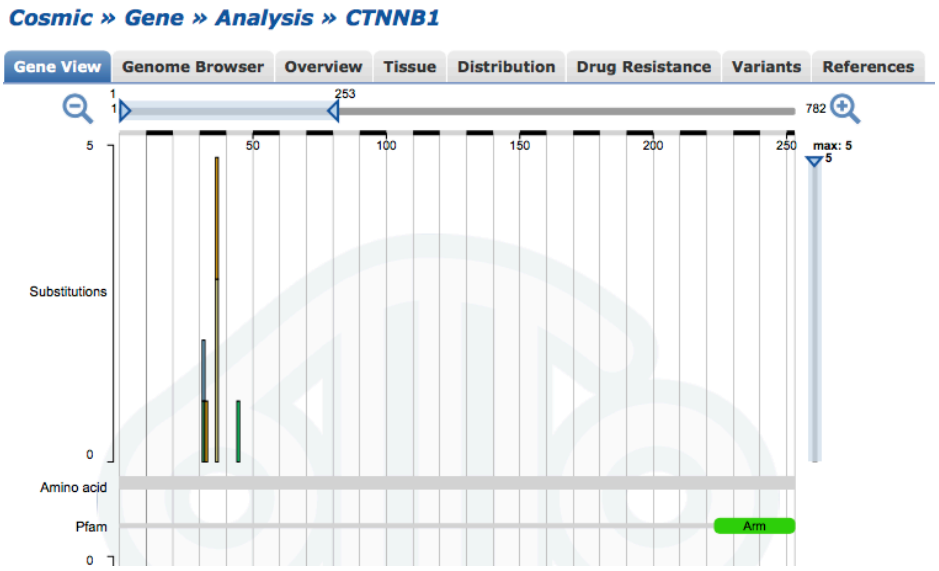



**Figure 2.1: COSMIC gene-browser overview of mutational burden in *CTNNB1***

TATCATTCTGCTTTTCTTGGCTGTCTTTCAGATTTGACTTTATTTCTAAAAAT
ATTTCAATGGGTCATATCACAGATTCTTTTTTTTTAAATTAAAGTAACATTTC
CAATCTACTAATGCTAATACTGTTTCGTATTTATAG**CTGATTTGATGGAGTTG**
**GACATGGCCATGGAACCAGACAGAAAAGCGGCTGTTAGTCACTGGCAGCAACA**
**GTCTTACCTGGACTCTGGAATCCATTCTGGTGCCACTACCACAGCTCCTTCTC**
**TGAGTGGTAAAG[GCAATCCTGAGGAAGAGGATGTGGATACCTCCCAAGTCCT**
**GTATGAGTGGGAACAGGGATTTTCTCAGTCCTTCACTCAAGAACAAGTAGCTG**
**]GTAAGAGTATTATTTTTCATTGCCTTACTGAAAGTCAGAATGCAGTTTTGAG**
**AACTAAAAAGTTAGTGTATAATAGTTTAAATAAAATGTTGTGGTGAAGAAAAG**
**AGAGTAA**

**Figure 2.2: *CTNNB1* region spanning chr3:41265875-41266356.** Bold letters are exonic regions. The region of interest is enclosed in square brackets. The primers targeting this region are highlighted in pink.

### 2.4.3.2 *FGFR3* primer design

FGFR3 mutations account for 47% of patients with bladder carcinoma (urothelial carcinoma) curated in the COSMIC database have mutations in FGFR3 out of a sample of 4082. 99.8% of these SNVs fall within 3 hotspot regions, Figure 2.3 and Table 2.2. Therefore primers were designed targeting, these regions alone, Figure 2.4-Figure 2.6.

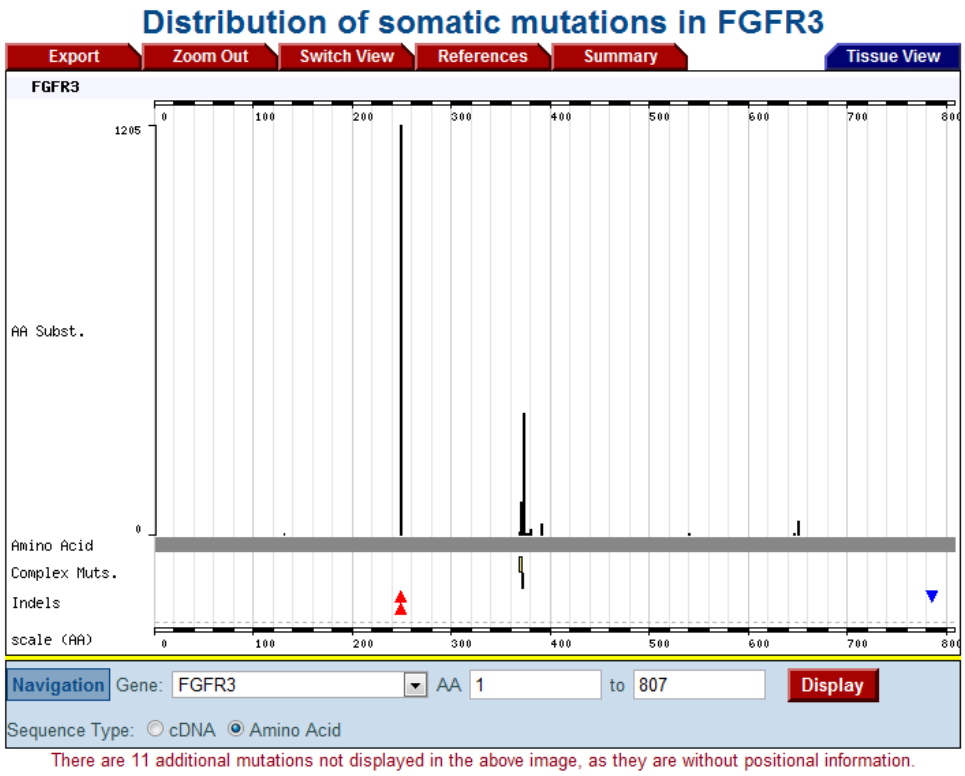



**Figure 2.3: Distribution of somatic mutations of *FGFR3 (Forbes et al.)***

TGGACGTGCTGGGTGAGGGCCCTGGGGCGGCGCGGGGGTGGGGGCGGCAGTGGCGGTGGTGGT

GAGGGAGGGGGTGGCCCCTGAGCGTCATCTGCCCCCACAG[AGCGCTCCC]CGCACCGGCCCA

TCCTGCAGGCGGGGCTGCCGGCCAACCAGACGGCGGTGCTGGGCAGCGACGTGGAGTTCCACT

GCAAGGTGTACAGTGACGCACAGCCCCACATCCAGTGGCTCAAGCACGTGGAGGTGAAtGGCA

GCAAGGTGGGCCCGGACGGCACACCCTACGTTACCGTGCTCAAGGTGGGCCACCGTGTGCACG

TGGGTGCCGCCGCTGGGGCTCCTGGGCTGGCCCCAAGGGTGCCCCTTGGCTGcGGGTTGCGTG

AGGATTTGGGTCTAGGGGTTGGAGCTTCGGGGGCAGAAGCTGTGGGGGTGCTTGTGGGGC

**Figure 2.4: DNA Sequence for Exon 6 of *FGFR3*.** The formatting of the figure is similar to **Figure 2.2**. hotspot SNVs are highlighted in red.

GGAGCCCCGTGGgGGGGGGGGCCAGGCCAGGCCTCAAcGCCCATGTCTTTGCAGCCGAGGAGG

AGCTGGTGGAGGCTGACGAG[GCGGGCAGTGTGTATGCAGGCATCCTCAGCTACGGGGTGGGC

TTCTTCCTGTTCATCCTGGTGGTGGCGG]CTGTGACGCTCTGCCGCCTGCGCAGCCCCCCCAA

GAAAGGCCTGGGCTCCCCCACCGTGCACAAGATCTCCCGCTTCCCGCTCAAGCGACAGGTAAC

AGAAAGTAGATACCAGGTTCTG

**Figure 2.5: DNA Sequence for Exon 8 of *FGFR3.*** The formatting of the figure is similar to **Figure 2.4**.

GGCCCTGCCCTGAGATGCTGGGAGCAGCGGGGAGAGGTGGAGAGGCTTCAGCCCTGCCTCCCA
CCCCTTCCCCAG**TGCATCCACAGGGACCTGGCTGCCCGCAATGTGCTGGTGAC**CGAGGACAAC
**GTGATGAAGA**TCGCAGACTTCGGGCTGGCCCGGGACGTG**[CACAACCTC**GACTACTACAAGAA
**GACgACCAAC]**GTGAGCCCGGCCCTgGGGTGCgGGGGTGGGGGTCATGCCAGTAGGACGCCTG
GCGCCAACACCGCCTTCCCACA

**Figure 2.6: DNA Sequence for Exon 13 of *FGFR3.*** The formatting of the figure is similar to **Figure 2.4**.

### 2.4.3.3 *NFE2L2* primer design

Nuclear factor (erythroid-derived 2)-like 2 is a basic leucine zipper protein that is involved in the regulation of antioxidant expression. In 2% of patients with BC, *NFE2L2* mutations exist. SNVs in BC, have only been reported in one region but are reported in only 2 hotspots in all cancer types, Figure 2.7. Therefore, 3 primer pairs were designed, targeting both regions.

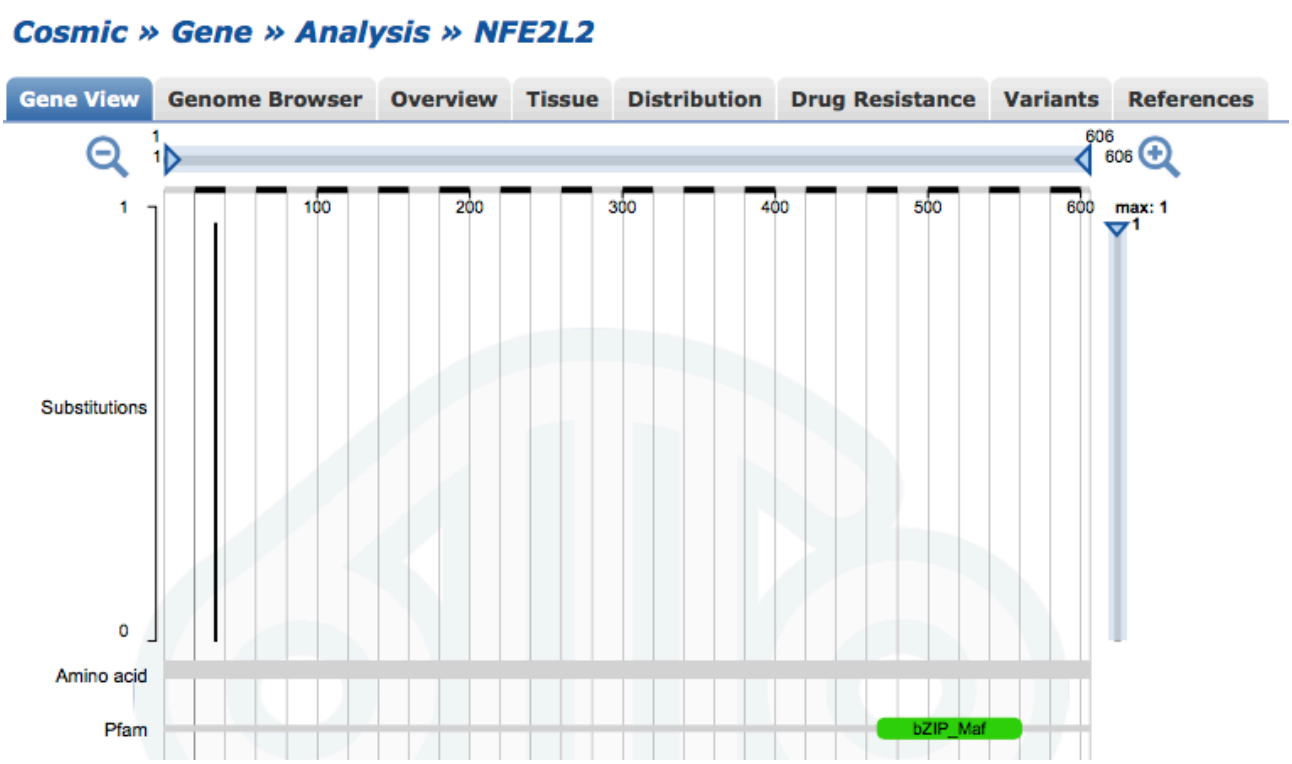



**Figure 2.7: Distribution of somatic mutations of *NFE2L2 (Forbes et al.)***

CTCAGTGTTTCCTTAAACCTGCCATAACTTTCCCAAGAACTGAGTACTCTGTA
CCTGGGAGTAGTTGGCAGATCCACTGGTTTCTGACTGGATGTGCTGGGCTGGC
TGAATTGGGAG[AAATTCACCTGTCTCTTCATCTAGTTGTAACTGAGCGA]AA
AAGGCTTTCTCTTGCTCCTTTTGGAGTTGTTCTTGTCTTTCCTTTTCAAGTTT
TTTCTGTTTTTCCAGCTCATACTCTTTC[CGTCGCTGACTGAAGTCAAATACT
TCTCGACTTACTCCAAG]A[TCTATATCTTGCCTCCAAAGTATGTCAATCA]A
ATCCATGTCCTATGTTTAAGACAAAAAAAGGAAGGAGAGAGCTCATGTTTTTT
AAATGACAGATACCACATAAATTAATTCACATTATGCCACTGTTGATGGTGGG

**Figure 2.8: DNA sequence for *NFE2L2* exon on chr2:178098679-178160622.** In this figure forward and reverse paired primers are highlighted in the same colour, whilst hotspot mutations curated by COSMIC are underlined.

### 2.4.3.4 *TERT* primer design

Approximately 70% of patients with bladder carcinoma have *TERT* mutations in their promoter region (Hurst *et al.;* Kinde *et al.;* Forbes *et al.*) 2 hotspots account for 77% of these mutations, with a further promoter region hotspot accounting for an additional 3% of mutations, Figure 2.9 and Table 2.2. Therefore, primer pairs for these regions were designed.

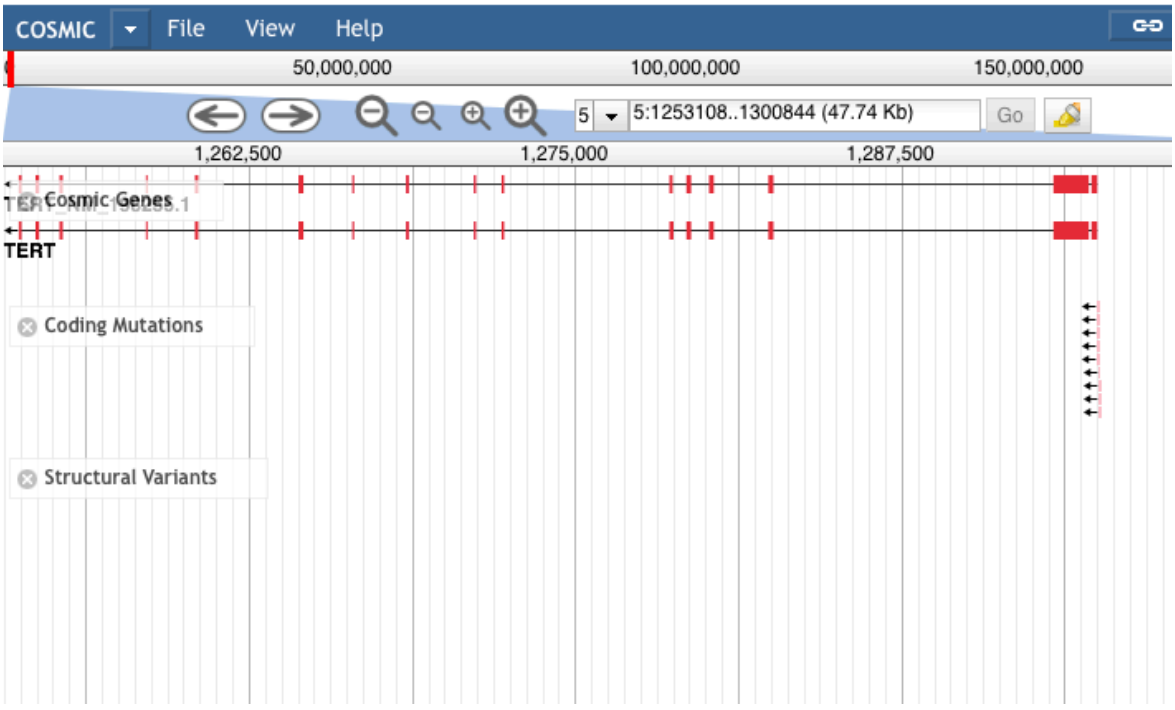



**Figure 2.9: Distribution of somatic mutations for *TERT* (Forbes *et al.*).** SNVSs are found in the promoter region of *TERT*.

**TCCCCGCGCTGCACCAGCCGCCAGCCCTGGGGCCCCAGGCGCCGCACGAACGTGGCCA**
**GCGGCAGCACCTCGCGGTAGTGGCTGCGCAGCAGGGAGCGCACGGCTCGGCAGCGGGG**
**AGCGCGCGGCATCGCGGGGGTGGCCGGGGCCAGGGCTTCCCACGTGCGCAGCAGG[AC**
**GCAGCGCTGCCTGAAACTC**GCGCCGCGAGGAGAGGGCGGGGCCG**CGGAAAGGAAGGGG**
**AGGG**GCTGGGAGGGCCCGGAGGGGGCTGGGCCGGGGACCCGGGA]GGGGTCGGGACGG
GGCGGGGTCCGCGCGGA**GGAGGCGGAGCTGGAAG**GTGAAGGGGCAGGACGGGTGCCCG
GGTCCCCAGTCCCTCCGCCACGTGGGAAGCGCGGTCCTGGGCGTCTGTGCCC

**Figure 2.10: DNA sequence for *TERT* promoter regions on chr5:1294977-1296504.** Major hotspot mutations curated by COSMIC are highlighted in read and minor hotspot regions (i.e. with a mutation rate of <3%) are highlighted in blue). TERT_1forward and reverse primers are highlighted in green and TERT_2 forward and reverse primers are highlighted in pink.

### 2.4.4 Primer Validation

Primer specificity for the target region was determined by conducting BLAT (Barbieri *et al.*) and in-silico PCR using UCSC genome browser (Kent *et al.*, 2002). Primers were further validated by PCR with genomic DNA extracted from the white blood cells of healthy male volunteers (Promega, PR-G1471). The PCR mastermix concentrations used for testing are listed in Table 2.4 and cycling parameters are listed in Table 2.5. Both the PCR reagent concentrations and the cycling conditions were designed to mimic TAm-Seq conditions as previously described (Forshew *et al.*). PCR products were checked against expected size using the Broad spectrum DNA kit on Bioanalyzer 2100 (Agilent G2938) or e-Gel Agarose Gel Electrophoresis System (Invitrogen).

| | Per 10µL reaction | Final conc. |
|---|---|---|
| 10X FastStart High Fidelity Reaction Buffer without MgCl2 (Roche) | 1 | 1X |
| 25 mM MgCl2 (Roche) | 1.8 | 4.5mM |
| DMSO (Roche) | 0.5 | 5% |
| 10mM PCR Grade Nucleotide Mix (Roche) | 0.2 | 200 µM ea |
| 5 U/µL FastStart High Fidelity Enzyme Blend (Roche) | 0.1 | 0.05U/µL |
| 20X Access Array Loading Reagent (Fluidigm) | 0.5 | 1X |
| 50 ng/µL Genomic DNA | 1 | 10ng/µL |
| Primer Pair (250nM) | 2 | |
| PCR Certified Water (Teknova) | 2.9 | |

**Table 2.4: Primer testing PCR master mix concentrations**

| PCR Stages | Number of cycles |
|---|---|
| 50ºC 2 minutes<br>70ºC 20 minutes<br>95ºC 10 minutes | 1<br>1<br>1 |
| 95ºC 15 seconds<br>60ºC 30 seconds<br>72ºC 1 minute | 10 |
| 95ºC 15 seconds<br>80ºC 30 seconds<br>60ºC 30 seconds<br>72ºC 1 minute | 2 |
| 95ºC 15 seconds<br>60ºC 30 seconds<br>72º 1 minute | 8 |
| 95ºC 15 seconds<br>80ºC 30 seconds<br>60ºC 30 seconds<br>72ºC 1 minute | 2 |
| 95ºC 15 seconds<br>60ºC 30 seconds<br>72ºC 1 minute | 8 |
| 95ºC 15 seconds<br>80ºC 30 seconds<br>60ºC 30 seconds<br>72ºC 1 minute | 5 |

**Table 2.5: PCR thermocycling conditions for primer testing.**

### 2.4.5 Identifying mutDNA through TAm-Seq

TAm-Seq was performed on cfDNA extracted from urine samples from patients with metastatic bladder cancer. Urine samples were collected during routine oncological follow up and were selected based on their high burden of disease. Ethical approval for collection was under Diamond (LREC 03/018) and approval for 5-6 patients was granted by the Cambridge Urological Biobanking board. Urine samples were processed according to Table 2.1 with exception that only 5ml were processed for each category.

DNA was extracted from urine supernatants using the modified Qiagen urine protocol. Samples were quantified using digital PCR, as described in Chapter 2.4.2, with the 97bp *RPP30* probe and 50ng, where possible, was utilised for pre-amplification in the TAm-Seq work flow.

The TAm-Seq process was performed according to the previously published protocol (Forshew *et al.*, 2012). In brief, it involved a limited cycle PCR (pre-amplification) of each sample. Each sample was multiplexed with a pool of all bladder panel primers. The pre-amplified material was then cleaned using an exonuclease (EXOSAP TM) before diluting and undergoing target specific PCR using a microfluidic platform (Fluidigm Access Array). The Access array platform allows each sample to be split into 48 wells of ~35nl (Fluidigm, 2017) where each reaction chamber can undergo singleplex PCR with a single primer pair from the bladder panel. The amplified PCR material in each reaction chamber is then pooled so that material from each sample can be harvested from the access array. Samples are then barcoded in a further PCR reaction before pooling and cleaning to create a library suitable for submission for NGS.

NGS was performed using MiSeq (Illumina). Sequence alignment and clipping was performed by Mohammed Murtaza, James Morris and Francesco Marass using Samtools. Variant calling was performed using a custom designed calling pipeline, taking into account, the read depth, the number of abnormal reads and the level of background noise, reported presence of single nucleotide polymorphisms and presence of the same mutation in duplicate samples prior to calling SNVs.

## 2.5 Results

### 2.5.1 Urine DNA extraction

Only 5 out of the 7 healthy volunteers produced >150ml of urine for aliquoting according to Table 2.1. Volunteer A produced 80ml and only had multiple aliquots of 1 hour with EDTA processed. Volunteer D produced 60ml of urine and had a single 5ml aliquot for each processing variation. Volunteer D was therefore excluded from kit testing analysis. Volunteer E produced >250mls of urine allowing aliquots of multiple processing methods and multiple replicates. During the initial phase of testing, only urine samples that were frozen within 1 hour and had EDTA added were analysed, see Figure 2.11.

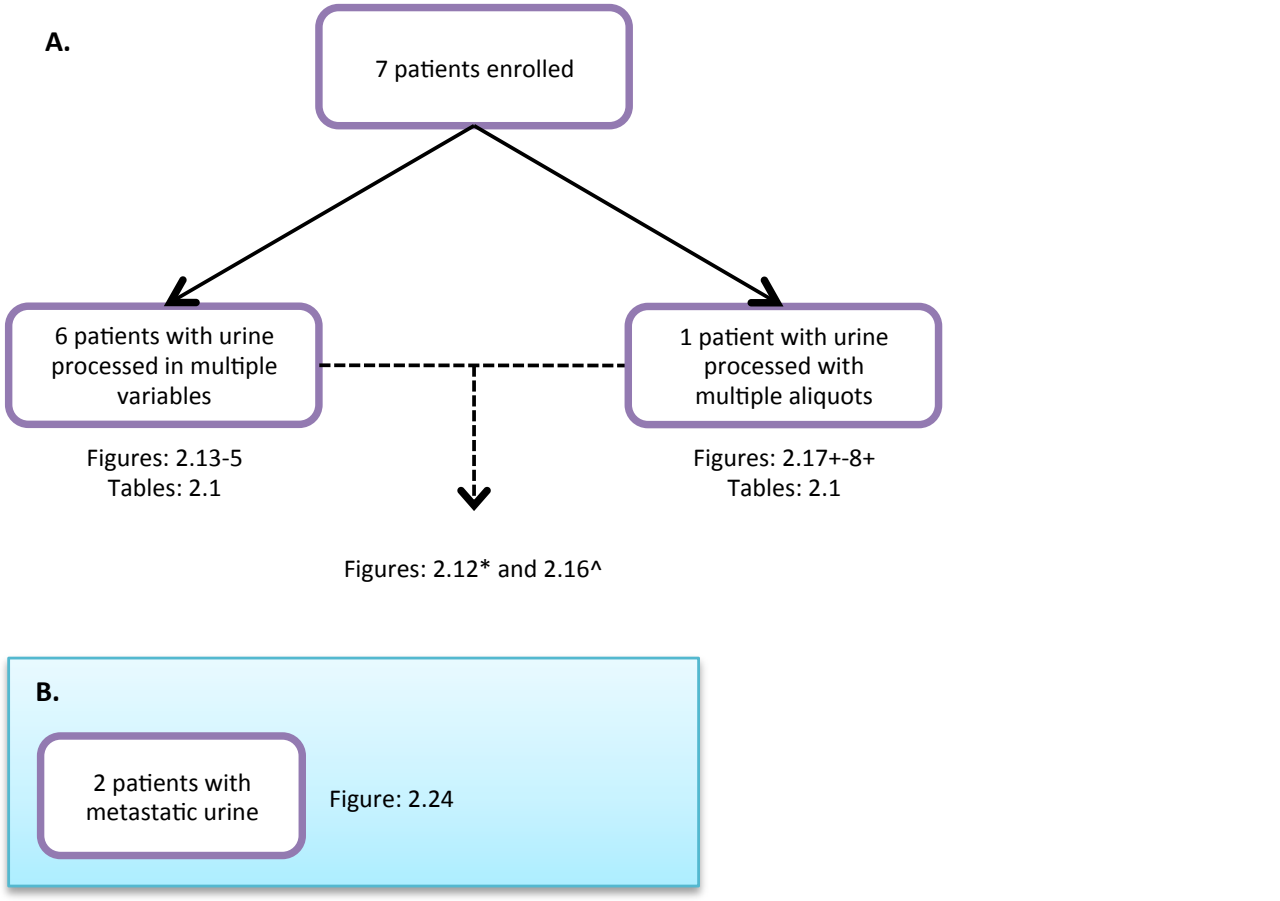



**Figure 2.11: Flow diagram for patient samples used to evaluate urinary DNA.** Numbers of patients analysed are included in purple boxes, with figures and tables listed adjacent. **A. Samples collected for urinary cfDNA extraction optimisation.** 7 healthy volunteers provided urine samples for proof of principle analysis to ensure results were reproducible across multiple volunteers. 2/7 volunteers were unable to provide enough urine samples for multiple processing affecting some analyses. *__Figure 2.12__ excluded urine from volunteer D due to lack of sample. ^__Figure 2.16__ excluded urine from volunteers A and D due to lack of sample. +__Figure 2.17__-8 analysed volunteers G and E replicate urine samples. **B. Samples collected for optimisation of mutDNA analysis.** Cambridge Urology Biobank approval was received to collect urine from 7 metastatic bladder cancer patients to allow pilot panel detection of SNVs in sufficient numbers of samples. Only 2 patients were recruited for before preliminary analysis of MIBC patient urine samples demonstrated detectable SNVs, see Chapter 3.6. Therefore further recruitment was stopped.

Part-way through this analysis, the SNOVA nucleic acid kit was withdrawn from production and SNOVA ceased to trade. However a new company, NeoGeneStar set up by the same directors released cfDNA extraction kits based on similar magnetic bead chemistry. These were included in analysis of DNA extraction for remaining samples. Figure 2.12 depicts the GE copies detected using dPCR with each of the 4 extraction kits. The urine samples used for this comparison were processed within 1 hour, with the addition of EDTA but were not centrifuged. Shapiro Wilk testing demonstrated a non-Gaussian distribution for samples extracted by Norgen (W= 0.8141, p= 0.0042) and Qiagen kits (W= 0.7414, p= 0.0005). Gaussian distributions were suggested for samples extracted by SnoMag (W= 0.9027, p= 0.3474) and NeoGeneStar kits (W= 0.9617, p= 0.82). Due to the comparison of data sets with mixed distributions, non-parametric analysis was used to determine statistical significance of differing yields. Kruskall Wallis demonstrated statistically significant differences for the yield obtained by the extraction kits tested. Dunn testing demonstrated statistically higher yield from both Qiagen and Norgen extraction when compared with SnoMag (p=0.0028 and p=0.0088). Total DNA yields and cfDNA concentrations for each sample are listed in Appendix A-1 and indicates that of the column-based beads the Qiagen nucleic acid kit has a higher yield for every case apart from Volunteer E. However, there was no statistically significant difference between urine extraction with Norgen, NeoGeneStar nor Qiagen kits (p= 0.3053 for Qiagen versus Norgen, p=0.3752 from NeoGeneStar versus Norgen and p= 0.2516 for NeoGeneStar versus Qiagen). Of note, extracted DNA from volunteer E was further aliquoted into two replicates for unspun urine and further dPCR performed, the low amplifiable copies/ml detected imply that extraction is likely to have failed in this sample.

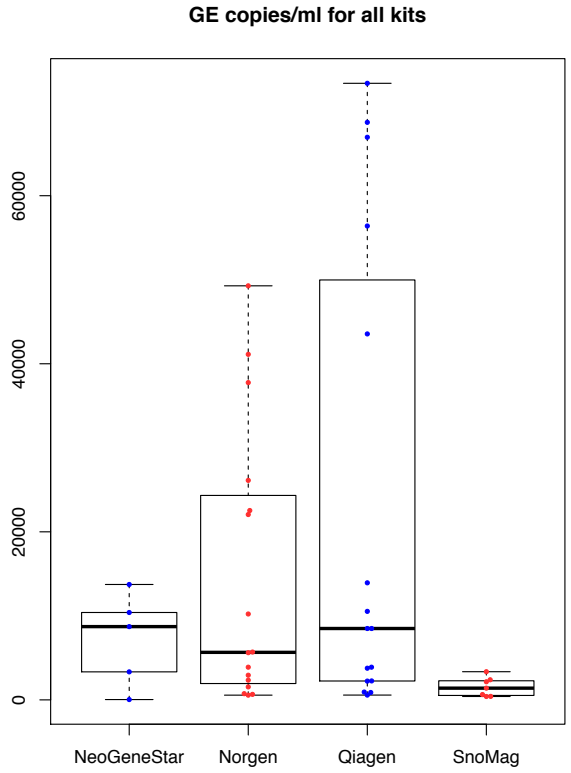



**Figure 2.12: Box plots suggesting that urine DNA extraction with Qiagen Circulating Nucleic Acid and Norgen slurry kits achieved the highest yield of DNA.**

### 2.5.2 Urinary DNA Processing

Based on these results, the Qiagen Circulating Nucleic Acid extraction kit was selected and subsequently used to extract remaining aliquots of urine samples processed according to Table 2.1. This was due to the Qiagen kit having the highest mean cfDNA GE copies/ml isolated during the comparison and, it having been extensively used by the Rosenfeld lab to extract cfDNA from plasma samples. Use of the Qiagen extraction kits also facilitated further comparison of matched plasma and urine samples. DNA yields and cfDNA concentrations are listed in Appendix A-2. Based on this work, this kit has since been in use in the lab for additional projects where urine has been collected and analysed. At a later stage, a new method of automating the Qiagen extraction through the qiasymphony arose and we compared manual and automated kits in section 3.6.2.

Urine samples that were subjected to centrifugation had lower GE copies/ml than those that were not (Figure 2.13), The centrifuged urine samples had a mean of 2,050 GE copies/ml of DNA extracted, whilst un-centrifuged samples had a mean of 33,060 GE copies/ml (Kruskall-Wallis p = 1.47xe$^{-5}$).

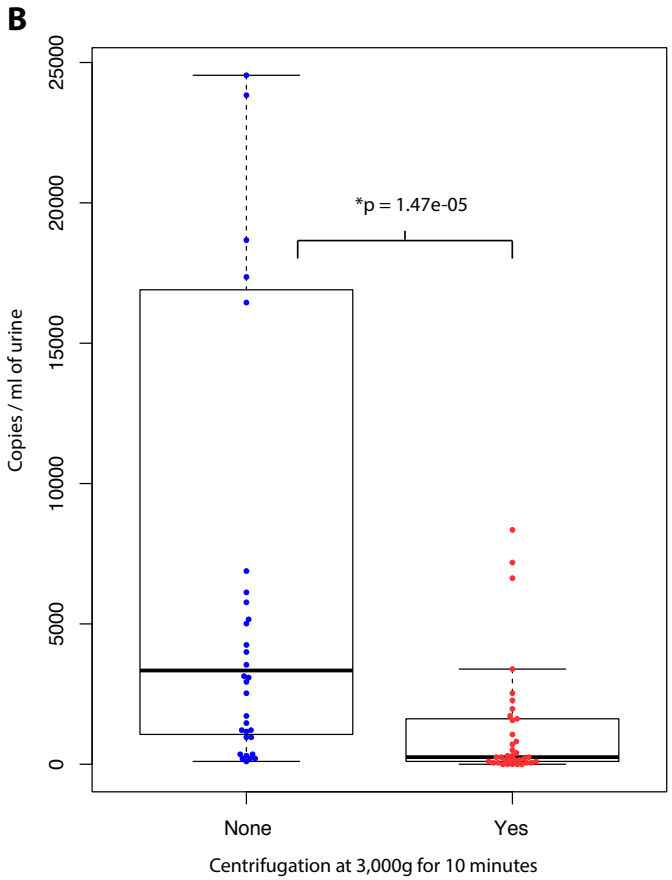



**Figure 2.13: Cell-free DNA levels are higher in urine samples processed without a centrifugation step.**

The addition of 10mMol EDTA to urine samples resulted in a higher concentrations of cfDNA when compared to samples that did not have EDTA (mean of 19,350 GE copies/ml versus 15,170 GE copies/ml respectively). However, this was not a statistically significant difference, Kruskal Wallis p = 0.0809.

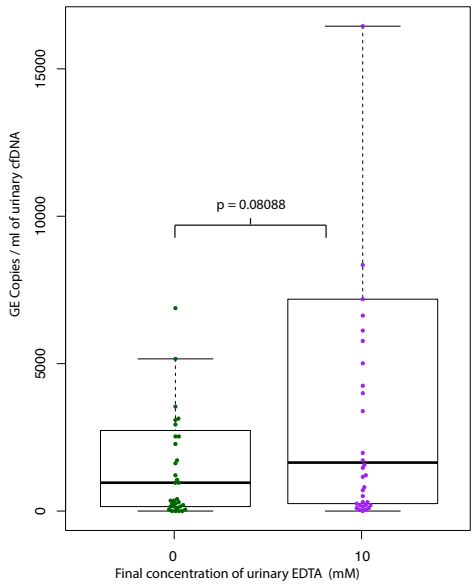



**Figure 2.14: 10mMol EDTA may improve cell-free DNA yield**

Reducing the time taken to process urine samples resulted in increased DNA yield. Samples processed at 1 hour had a mean GE copies/ml of 23,370 compared to 22,570 GE copies/ml for samples processed at 6 hours and 5,083 GE copies/ml for urine samples processed at 48 hours, Figure 2.15A. Figure 2.15B depicts the change in GE copies/ml from the matched 1 hour specimen. For samples processed at 6 hours 11/20 had lower DNA yield than when processed at 1 hour. For samples processed after 48 hours, 14/20 had lower DNA yields than those processed at 1 hour suggesting that processing samples earlier improves the yield of DNA. This difference was statistically significant (p= 0.0121, Friedman test), Figure 2.15A. For urinary mutDNA to be clinically useful, longitudinal sampling would ideally be analysed, typically when patients are attending clinic. We therefore attempted to limit the time taken to process urine samples to 6 hours, so that it would be feasible for research nurses to collect all samples from a clinic before processing them.

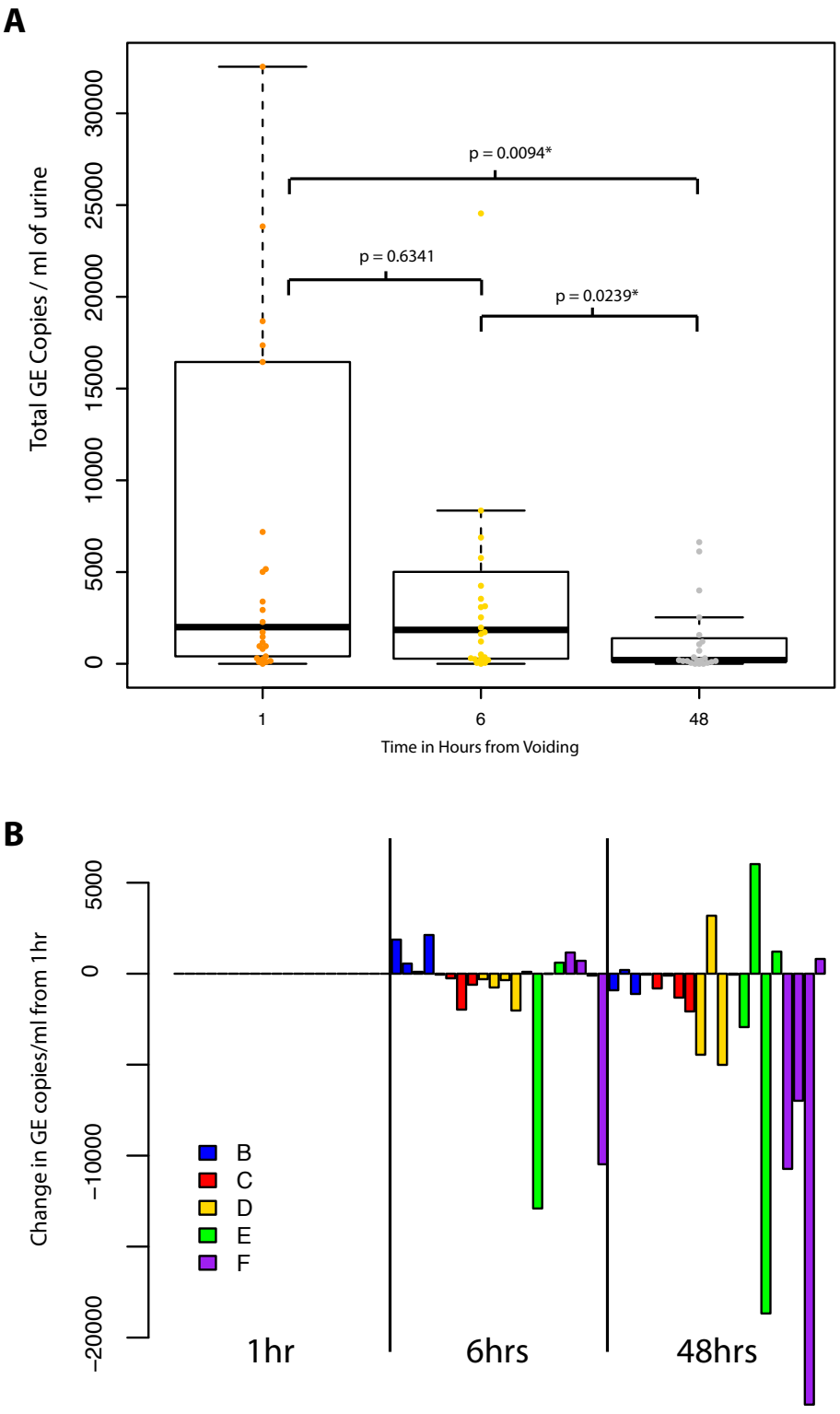



**Figure 2.15: Reduced time to processing of fresh urine improved cfDNA yield. A. Total GE copies/ml for urine samples extracted at 1, 6 and 48 hours.** Shapiro Wilk testing revealed non-Gaussian distribution for samples processed at 1 hour (W= 0.4004, p= 2.88xe$^{-09}$) for samples processed at 6 hours (W= 0.3637, p= 3.39xe$^{-09}$) and for samples processed at 48 hours (W= 0.2794, p = 1.18xe$^{-09}$). Friedman testing revealed statistically significant differences between the cfDNA amounts obtained with different processing times (p=0.0121). **B. Change in GE copies/ml (y axis) for matched samples processed at 1, 6 and 48 hours (x axis).** GE copies/ml at 1 hour were subtracted from GE copies/ml for matched samples at 6 and 48 hours. Samples are coloured by patient.

### 2.5.3 Urinary DNA Size

To investigate whether there may be longer DNA fragments in urinary cfDNA, an RRP30 dPCR assay targeting an amplicon of 360bp was used to quantify DNA eluted from patient G. Centrifuged and non-centrifuged urine samples that were processed within 1 hour and had EDTA added were assessed. Initial results suggested a drastic reduction in 360bp fragments. Comparison between centrifuged and un-centrifuged samples revealed a respective mean of 1,058 and 3,941 GE copies/ml for amplicons ≥360bp, see Figure 2.16, indicating that longer urine cfDNA fragments can be found in un-centrifuged urine. Indeed, this would be expected as not performing centrifugation would retain cellular debris in the urine specimen. Subsequent lysis during the extraction protocol would increase the proportion of longer cellular DNA fragments. However, this finding was not statistically conclusive (p= 0.0633) possibly due to the small sample size (n=21).

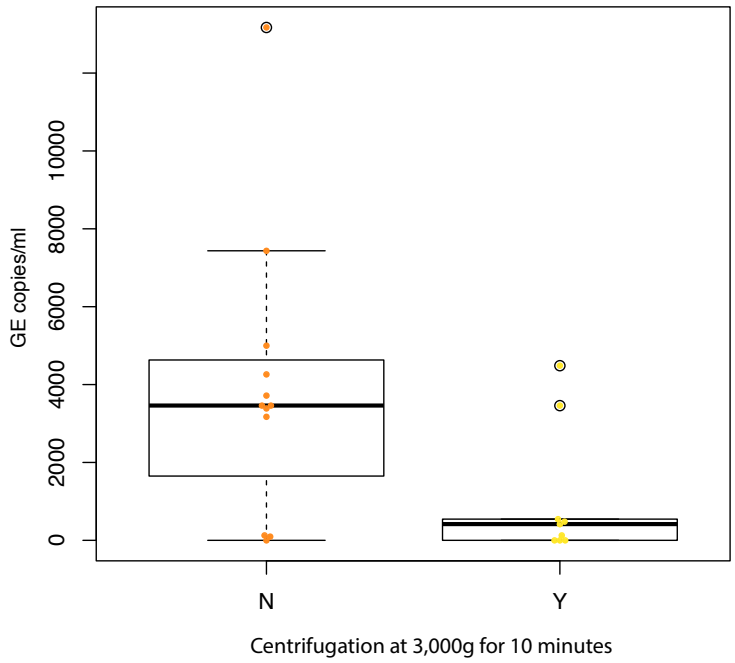



**Figure 2.16: GE 360bp (and larger) copies/ml for samples of urine that were centrifuged (n=9) and those that were not (n=12).**

Figure 2.17 demonstrates that smaller sized primer pairs amplify more fragments of cfDNA. This is likely due to the smaller primer pairs amplifying both small and large fragments of DNA. However, even after subtraction of longer fragments, i.e. fragments the 360bp assay detected, there are numerous additional amplifiable copies that have been amplified by the 65bp and 97bp assays. This implies that there are fewer long urinary cfDNA fragments than short fragments. Kruskal-Wallis testing confirms a statistical

significance in this difference with 360bp urine fragments being shown to be significantly lower than 65bp and 97bp fragments (p= 0.0029 and p= 0.0156 respectively).

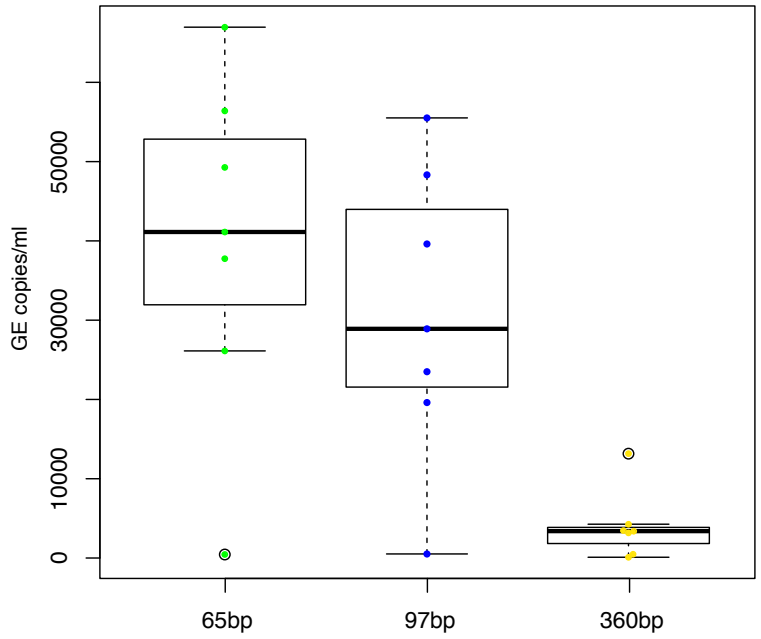



**Figure 2.17: Comparison of GE copies/ml for urine cfDNA amplicons of 65bp, 97bp and 360bp lengths from 7 replicate urine aliquots from 2 patients.** Each sample was processed with EDTA and frozen within 1hour. PCR reactions were carried out in singleplex, to avoid bias from preferential amplification of amplicons of different lengths. Kruskal-Wallis testing demonstrated a significant difference between the 3 sizes (p=0.0150). Post-hoc Dunn testing revealed significant differences less 360bp fragments when compared with 65bp and 97bp DNA fragments (p= 0.0029 and p= 0.0156). There was no difference between 65bp and 97bp size fragments in this small sample-set.

Figure 2.18 demonstrates no statistically significant difference between 65bp and 97bp fragment sizes, though the median GE copies/ml for 65bp fragments was higher than for 97bp fragments (28,910 versus 41,120 GE copies/ml). The small sample size and 32bp difference between sizes tested may have impacted on the ability to detect a difference. Tsui *et al.* demonstrate that most foetal urinary cfDNA lies between 29bp and 45bp in length and most maternal urinary cfDNA between 29bp and 85bp (Tsui *et al.*, 2012). A considerably higher proportion of 97bp amplicons were amplified from our urine samples than expected based on the data of Tsui *et al.* alone and this is likely to be due to differing urine processing (urine centrifugation and filtration by Tsui *et al.* versus centrifugation alone) and DNA extraction techniques (Promega Wizard Plus by Tsui *et al.* versus QIAmp Circulating Nucleic Acid kit) used. However, both data sets emphasise the importance of targeting short urine cfDNA fragments and therefore, primers were designed to target amplicons of 60bp -100bp in size.

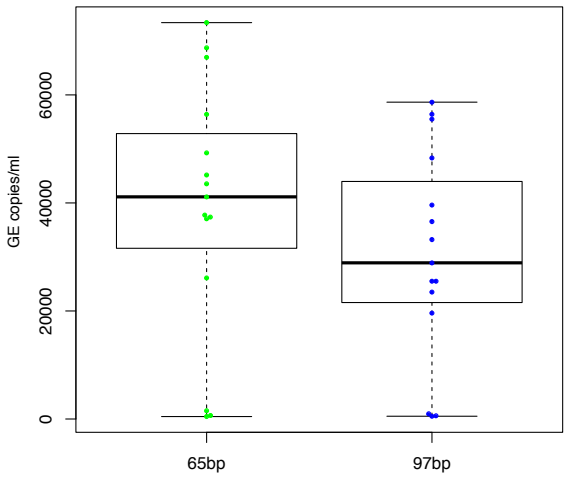



**Figure 2.18: Comparison of GE copies/ml for urine cfDNA amplicons of 65bp and 97bp from 15 replicate aliquots from 2 patients.** dPCR reactions were carried out in singleplex. Kruskal-Wallis testing showed no difference (p= 0.1776).

### 2.5.4 Primer Validation

Bioanalyser traces revealed no PCR product bands for negative control wells and tight bands of PCR products at the expected amplicon size range for each of the *FGFR3* primer pairs investigated (Figure 2.19).

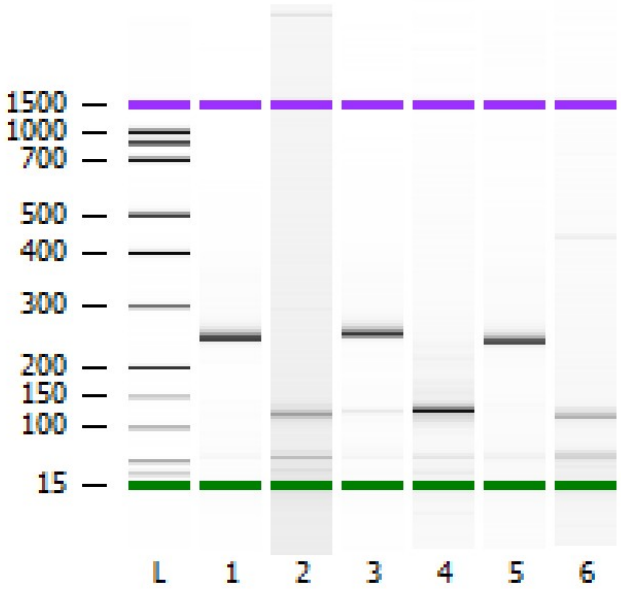



**Figure 2.19: Bioanalyser trace of PCR products with 3 *FGFR3* primer pairs.** X-axis key for primer pairs tested in each lane: 1:FGFR_Ex6 region, 2: negative PCR control (FGFR_Ex6 region without DNA), 3: FGFR_Ex8 region, 4: negative PCR control (FGFR_Ex8 region without DNA), 5: FGFR_Ex13 region, 6: negative PCR control (FGFR_Ex13 region without DNA).

Primer pairs targeting *TERT*, had no or very weak PCR product bands in the expect size range and multiple weak bands in other size ranges (Figure 2.20), this result was replicated on repeat experimentation (data not shown).

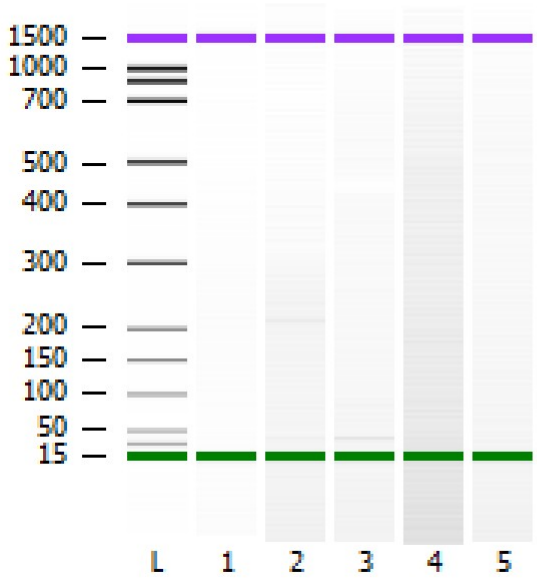



**Figure 2.20: Bioanalyser trace of PCR products with 3 *TERT* primer pairs.** X-axis key for primer pairs tested in each lane: 1:TERT1_F - TERT1_R, 2: negative PCR control (TERT region 1), 2: TERT_2_F - TERT_2_R, 4: negative PCR control (TERT region 2), 5: positive PCR controls (using FGFR3_Ex6 primers).

Bioanalyser testing revealed no PCR product bands for the negative control well and a tight band of PCR product at the estimated size range for the positive control band. Tight product bands, at the expected size, were present for all 3 primer pairs targeting *NFE2L2* and the supplementary primer pair targeting *CTNNB1*, see Figure 2.21.

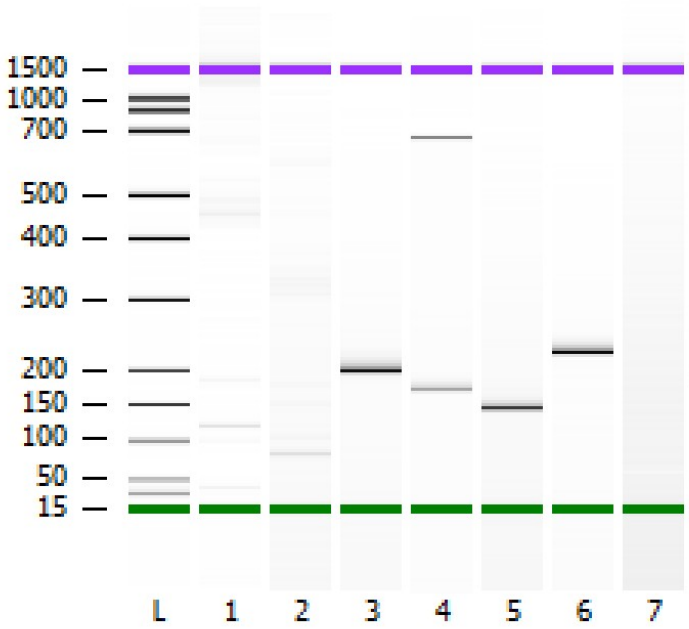



**Figure 2.21: Bioanalyser trace of PCR products with 3 *NFE2L2*, 1 *CTNNB1* and 2 re-designed *TERT* primer pairs.** X-axis key for primer pairs tested in each lane: L: DNA ladder, 1: TERT region 1F-2R, 2: TERT region 2BF – 2R, 3: NFE2l2 region 1 primers, 4: NFE2l2 region 2, 5: NFE2l2 region 3, 6: CTNNB_SUPP, 7: negative control for PCR (CTNNB_SUPP without DNA).

#### 2.5.4.1.1 *TERT* re-design

Due to the failure of the primer pairs to specifically bind to the *TERT* promoter region, multiple new primer pairs were designed, focusing on the most commonly mutated hotspot. Multiple forward primers were designed to be tested with one or two reverse primers, see Figure 2.22 and Table 2.6. In addition, the previously designed forward primer, TERT_1F was tested with the primer TERT_2R with the aim of avoiding repetitive regions, but creating an amplicon with ~250bp.

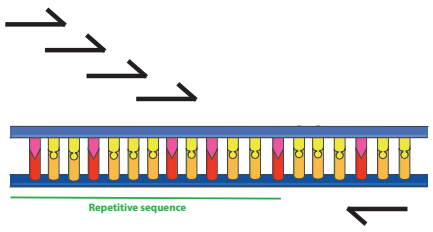



**Figure 2.22: Alternative strategy to design primer pairs for *TERT* promoter region.** Multiple forward primers in the repeat region were designed to pair with one or two reverse primers in a relatively heterogeneous region. Testing of multiple primer pairs was performed in singleplex.

| Name | Seq | Tm | Len | SNPs | Insilico PCR |
|---|---|---|---|---|---|
| TERT_Prom_2B_F | CGGAAAGGAAGGGGAGGG | 59.0 | 18 | No | OK |
| TERT_Prom_2B_R | CTGCCCCTTCACCTTCCAG | 60 | 19 | No | OK |
| TERT_Prom_2C_F | GGCCGCGGAAAGGAAG | 57.9 | 16 | No | OK |
| TERT_Prom_2C_R | CGTCCTGCCCCTTCACC | | 17 | No | OK |
| TERT_Prom_2D_F | CTGCCTGAAACTCGCGC | | 17 | No | OK |
| TERT_Prom_2E_F | CTGAAACTCGCGCCGC | | 16 | No | OK |
| TERT_Prom_2F_F | AGGAAGGGGAGGGGCT | | 16 | No | OK |
| TERT_Prom_2G_F | CTTCCCACGTGCGCAG | | 16 | No | OK |

**Table 2.6: Alternative designs for TERT promoter region primer pairs.**

Figure 2.21 demonstrates that re-designed TERT primer pairs were not able to produce a specific PCR product within a tight expected amplicon size range despite, whilst appropriate negative and positive controls worked gave expected results. Similarly, Figure 2.23 also shows poor specificity for re-designed primer pairs targeting the promoter region.

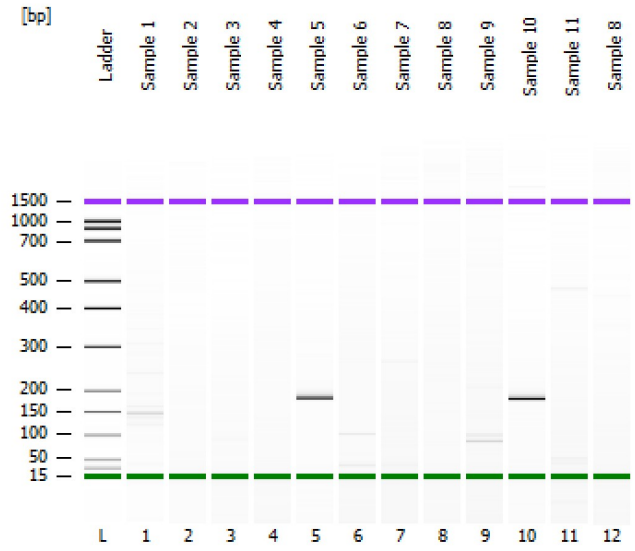



**Figure 2.23: Bioanalyser trace of PCR products with re-designed *TERT* primer pairs.** X-axis key for primer pairs tested in each lane: 1: TERT_Prom_2C_F - TERT_Prom_2C_R, 2: TERT_Prom_2G_F - TERT_Prom_2B_R, 3: TERT_Prom_T2D_F - TERT_Prom_2B_R, 4:TERT_Prom_2E_F - TERT_Prom_2B_R, 5: FGFR3 Exon 13, 6: negative PCR control (FGFR3 Exon 13 without DNA), 7: TERT_Prom_T2C_F - TERT_Prom_2B_R, 8: Prom1R-T2BRv, 9: TERT_Prom_T2_F - TERT_Prom_2B_R, 10: FGFR3 Exon 13, 11: negative PCR control (FGFR3 Exon 13 without DNA).

A list of final primers used in the pilot bladder cancer panel is included in Appendix A-3.

### 2.5.5 Mutant DNA analysis from urine

Only 2 patients were enrolled in this arm of the study. Both patients were able to provide >120ml of urine. Each urine sample was processed and aliquots were frozen according to Table 2.1. Digital PCR was performed, on DNA extracted from the urine supernatants.

NGS with 150bp paired end using 2 lanes of a MiSeq (Illumina) instrument resulted in approximately 50 million reads with the majority of reads have a mean sequence quality (PHRED score) of above 30, see Figure 2.24A. Primer dimer removal, quality clipping of the reads followed by alignment of the reads to the human reference genome resulted in approximately 35 million useable reads, Figure 2.24B.

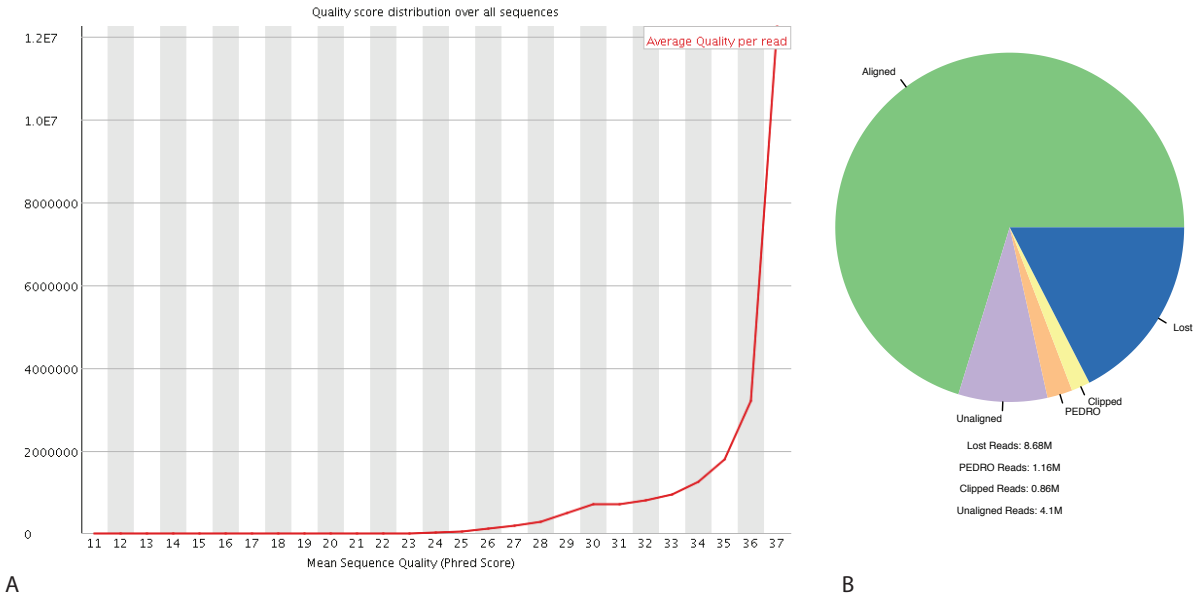



**Figure 2.24: Basic statistics for pooled library submission to MiSeq (Illumina) NGS. A.** PHRED score CDF graph demonstrating the high quality scoring for most of the reads. **B.** Pie chart showing that most reads were useable.

The median coverage was 5501 reads per base across all samples. Barcodes representing uniquely positioned No Template Controls (NTCs) were easily deducible by their low read count (median read count of <2, compared with median read count of 5823 in samples with patient / control DNA). These matched the position NTCs were inputted, indicating that sample switching was unlikely during the experiment. Despite other experimental metrics performing well, there were no SNVs that were detected that passed quality filtering. For this reason, the expected optimum processing method for mutant cfDNA purification was used.

## 2.6 Original Contributions to Knowledge

1. Urine DNA extraction by SNOVA kits gave a statistically significant lower yield than the other urine DNA extraction. It is likely that QIAmp circulating nucleic acid kit produced the highest yield, though this was not statistically significant.
2. Centrifugation of urine samples within 6 hours, likely with the immediate addition of EDTA, will likely improve cfDNA yield for mutDNA analysis.

Results from these investigations have been incorporated to form the Rosenfeld group urine extraction and processing SOPs and are being shared with international collaborators for future projects.

# CHAPTER 3: MUTANT DNA ANALYSIS IN MIBC

## 3.1 Synopsis

*(The following text has been excerpted from work published in Scientific Reports)*

Patients with Muscle Invasive Bladder Cancer (MIBC) have a poor 5-year overall survival (OS) rate. Though there is a 6% improvement in OS with Neo-Adjuvant Chemotherapy (NAC), many patients fail to respond. Analysis of mutant DNA in body fluids (mutDNA) could be used to allow non-invasive identification of NAC failure and tracking of clonal kinetics (in real-time) during systemic chemotherapy.

We prospectively collected 248 liquid-biopsy samples including plasma, cell pellets from urine (UCP) and urine supernatant (USN) from multiple time-points of 17 MIBC patients undergoing NAC. We assessed Single Nucleotide Variants (SNVs) and Copy Number Alterations (CNAs) in mutDNA using Tagged-AMplicon Sequencing (TAm-Seq) and shallow Whole-Genome Sequencing (sWGS).

MutDNA was detected in 35.3%, 47.1% and 52.9% of pre-NAC plasma, UCP and USN samples respectively. Though urine samples contained higher levels of mutDNA ($p=<0.001$), some mutations were only identified in plasma suggesting comprehensive analysis requires sampling multiple fluids. MutDNA was detected in 49/116 samples from patients who recurred, compared with 4/117 samples from non-recurrent patients ($p<0.001$) and SNVs were present at higher allele fractions ($p<0.001$). In 12 patients mutDNA was assessed 2-3 weeks after starting NAC, and persistence of mutDNA (with SNV allele fractions >0.5%) was associated with disease recurrence (83% sensitivity, 100% specificity, $p=0.003$). Longitudinal analysis of SNVs and CNAs in mutDNA demonstrated tumour evolution under the selective pressures of surgery and NAC in several cases. In one case, urine analysis tracked two distinct cancer clones with non-overlapping sets of *TP53* mutations.

This proof of principle study shows that mutDNA analysis may serve as an early biomarker for recurrence, offering MIBC patients an opportunity to switch NAC or expedite surgical resection in a timely manner. Furthermore, BC patients might benefit from tracking clonal kinetics during systemic therapy. These results warrant further evaluation of mutDNA in larger trials.

## 3.2 Publications arising from this work

*Work presented in this chapter is published in Scientific Reports. The text is therefore excerpted from this manuscript except to extend explanations, display updated data and provide more in-depth references.*

**KM Patel**[#], KE van der Vos[#], CG Smith[#], FC Mouliere, D Tsui, J Morris, D Chandrananda, F Marass, D van den Broek, DE Neal, VJ Gnanapragasam, T Forshew, BW van Rhijn, CE Massie, N Rosenfeld*, MS van der Heijden*. Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. Scientific Reports. 2017. 7:5554. http://dx.doi.org/10.1038/s41598-017-05623-3

[#]These authors contributed equally.

* These authors are joint senior supervisors

## 3.3 Aims

My primary objective was **to explore whether urinary mutDNA can be a clinically useful biomarker in bladder cancer**. Through international collaborations, I collected plasma, urine cell pellets and urine supernatant to investigate:

1. **Which body fluid, if any, allow for mutDNA analysis in MIBC:** I investigated mutDNA in fresh urine and plasma samples taken concurrently from MIBC patients undergoing NAC. These samples were used to compare both the number of times a mutation is detected and the AF of the mutations.
2. **The ability of mutDNA to predict response to NAC and outcome:** Each MIBC patient had sequential samples collected prior to each cycle of NAC. These samples were assessed to determine whether presence of mutDNA or mutDNA AF changes at early time points may be utilised to predict response to NAC.
3. **Whether tumour evolution can be monitored using mutDNA:** Use of a bladder panel, rather than patient specific probes may also allow monitoring of tumour clones. Sequential samples from MIBC patients undergoing NAC were used to assess whether tumour evolution occurs.

## 3.4 Methods

### 3.4.1 Sample Collection:

Approval according to Dutch national guidelines was obtained (N13KCM/CFMPB250). All patients gave informed consent to participate in this study. Formalin-fixed paraffin-embedded (FFPE) tumour blocks from the transurethral resection (TUR) samples were collected from referring hospitals. Slides were cut from FFPE blocks and used for Haematoxylin and Eosin (H&E) staining and DNA extraction. H&E stains were evaluated to identify areas with >50% tumour cells, from which DNA was extracted using the QIAamp DNA mini kit (Qiagen, Germany). Urine and blood samples were collected 1-2 hours prior to each chemotherapy session and were processed as follows; 10ml of peripheral blood was drawn into K2-EDTA haematology tubes and centrifuged at 380g for 20mins. The buffy coat layer was carefully transferred before the remaining plasma was aliquoted and spun at 20,000g for 10mins. Urine samples were processed by optimised methods described in section 2.5.2. In brief, urine samples were processed with the immediate addition of EDTA and subsequent centrifugation at 380g for 10mins before aliquoting the urine supernatant. The remaining UCP (Urine cell pellet) was re-suspended in 1ml of PBS. All peripheral fluids were processed within 6 hours and stored at -80°C.

DNA was extracted from 3 ml of urine supernatant using the optimal extraction system as investigated in Chapter 2: Section 2.5.1. In brief, we used the Qiagen Circulating Nucleic Acid kit to extract DNA from urine supernatant. DNA from TUR, BUF, UCP and plasma specimens was extracted using the QIAamp FFPE Tissue, DNeasy, DNA Mini Blood Mini and, Circulating Nucleic Acid kits (Qiagen) respectively. DNA extracted from BUF and UCP samples were subjected to mechanical shearing using either a Covaris S220 or LE220 (Covaris, USA). TUR was also sheared using the same protocol due to improved FFPE preservation methods leading to longer DNA fragment sizes (Greer *et al.*, 1991). Recommended parameters were used to shear fragments to an average fragment length of 140-180bp. Successful shearing was confirmed by running 1µl of DNA on a High-Sensitivity Bioanalyser gel (Agilent, USA).

### 3.4.2 TAm-Seq:

Tagged-AMplicon Sequencing (TAm-Seq) primers were designed to assess the mutation status for hotspots or entire coding regions of 8 genes (Table 3.1). Genes were

incorporated into the panel based on the frequency of mutation in WGS studies of BC and on exon length. Primer details are listed in Appendix A-3. TAm-Seq libraries were prepared as previously described (Forshew *et al.*). Libraries were sequenced using an Illumina HiSeq 2500 (Illumina, USA).

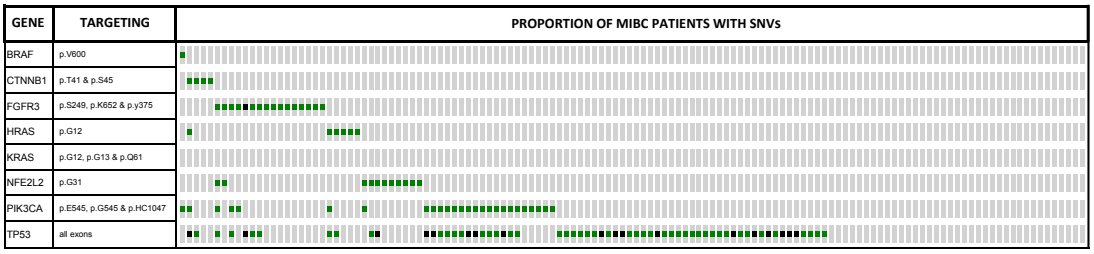



**Table 3.1: Genomic regions interrogated by TAm-Seq for SNV analysis.** ^Specificity of TERT assays was poor due to constraints of targeting short amplicons in the repetitive TERT promoter region. Data resulting from them were therefore excluded from downstream analysis. Other than TERT, 90-100% of mutations reported in the above listed genes were covered by the panel. Alterations in these genes would be expected to capture 72% of alterations reported in MIBC patients. *The prevalence of the mutations shown here are based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/.*, green squares represent missense mutations and black squares represent truncating mutations (Gao *et al.;* Cerami *et al.*).

### 3.4.3 Shallow Whole Genome Sequencing

Libraries were prepared from 10ng of plasma, USN, and TUR, BUF and UCP DNA using the ThruPLEX Plasma-Seq (Rubicon Genomics, USA) kit. Briefly, this involves end repair and 'A-tailing' of fragment ends. This precedes the ligation of truncated Illumina sequencer compatible adapters to fragment ends. Thermocycling of libraries completes the adapters through the addition of sample specific index sequences, and was performed as described in the Plasma-Seq protocol, using 8-14 (plasma, UCP and USN) or 8 (TUR and BUF) amplification cycles. Following amplification, libraries were cleaned with Agencourt AMPure XP beads (Beckman Coulter, USA) at a 1:1 (v/v) ratio and eluted in 30µl nuclease-free water. Successful library preparation was confirmed using a High-Sensitivity Bioanalyser gel and libraries were quantified using SYBR green based qPCR (Kapa Biosystems, USA). Libraries were pooled in an equimolar fashion and 125bp paired end sequencing was performed (to give a mean of 14.2 million reads per sample) using an Illumina HiSeq 2500 or HiSeq 4000.

### 3.4.4 Mutation Calling Criteria

TAm-Seq sequencing reads were aligned using Burrows-Wheeler Alignment (BWA) (Li and Durbin, 2009) and SNVs were detected using proprietary SNV calling software, the

principles of which were described previously (Forshew *et al.*). All mutation calling was performed blinded to the patient outcome. Patient-specific mutation calls were used to determine mutant AFs for any time-point from the 13 patients with detected SNVs, with thresholds defined by the highest of mutDNA AF >0.5% (technical threshold, see Results Section 3.6.5) or 1/GE copies inputted (sample threshold). Bases that contained a mutant call at a frequency below this threshold were not used for further analysis though AF's were retained for interpreting longitudinal mutDNA dynamics. NGS data were analysed using the R statistical software package (R Core Team 2015).

For sWGS analysis, sequence data was analysed using an 'in-house' pipeline developed by Dr. Christopher Smith (Postdoctoral scientist in the Rosenfeld group), adapted from published methods (Heitzer *et al.*). Briefly, this consisted of removing contaminant adapter sequences then aligning paired-end sequence reads to the human reference genome (GRCh37) using BWA (version 0.7.13) (Li and Durbin, 2009). SAMtools (version 1.3.1) (Li *et al.*, 2009) was then used to convert files to BAM format. PCR and optical duplicates were marked using Picard-Tools' (version 2.2.4) 'MarkDuplicates' feature and these were excluded from downstream analysis along with reads of low mapping quality and supplementary alignments.

CNA calling was performed in R (R Core Team 2015) using the QDNAseq pipeline (Scheinin *et al.*, 2014). Briefly, sequence reads were allocated into equally sized non-overlapping bins (1Mb and 50kb) throughout the length of the genome. Read counts in each bin were corrected to account for sequence GC content and mappability, and bins corresponding to previously 'blacklisted' (ENCODE) and manually blacklisted regions were excluded from downstream analysis. Within the QDNAseq package, bins were segmented using the 'Circular Binary Segmentation' algorithm (Seshan and Olshen, 2016) and significantly 'amplified' or 'lost' regions were called using CGHcall (van de Wiel *et al.*, 2007). CNAs were deemed present when a mutation was called in either 1Mb or 50Kb bin size to capture both large changes or focal changes in samples with low mutant:wild type AF. CNAs were called in peripheral fluids independent of the calls from the corresponding TUR sample.

We compared overall levels of copy number imbalance across the length of the genome in our samples by calculating a 'genome-wide imbalance score'. To generate this value, log2 adjusted read counts in a given 1Mb bin were compared against the equivalent value in a control sample. This control sample, which consisted of pooled sWGS data from 8 buffy-coat samples, was used for all pairwise comparisons. A linear model was fitted against all

autosomal bin values of the test sample vs. the control sample and the squared sum of residuals of this fit was calculated. To overcome inherent noise surrounding baseline (i.e. copy number neutral) in sWGS data, we only considered the sum of the 5% most extreme residual values to represent the 'genome-wide imbalance score'.

To indicate whether there was crossover between patients' samples from collection to the data interpretation stage, single nucleotide polymorphism (SNP) unique positions were identified from the 1000 genomes phase 3 per chromosome VCF files using Tabix (Li, 2011) and the amplicon start and end positions listed in Appendix A-3.

## 3.5 Statistical Inferences:

Statistical conclusions were impacted due to the proof of principle nature of the study. Where applicable we employed the following statistical analyses: To compare raw AF's of patients who did and did not recur, these two populations were plotted in the form of an empirical cumulative distribution function and assessed by applying the Kolmogorov–Smirnov test. Analyses of correlation between AFs at the 2$^{nd}$ NAC cycle and recurrence was performed using the "OIsurv" package in R (Diez, 2013). Survival curves were generated using the 'survfit' function, which uses the Logrank test to compare differences (in the presence of censoring) (Diez, 2013). Exact binomial confidence limits for sensitivity and specificity were calculated using the 'epiR' package in R (Stevenson, 2016). The SNV AFs in different sample types was compared using a Kruskal-Wallis rank sum test after testing for normality. sWGS profiles were compared by applying linear-regression modelling to the log ratios of two samples and adjusted $R^2$ values were generated using the linear model function in R.

## 3.6 Results

### 3.6.1 Patient recruitment for longitudinal analysis of mutDNA kinetics

Patients attending the Netherlands Cancer Institute (NKI) for cisplatin-based NAC were recruited between March 2014 and October 2015. We analysed 282 samples from 17 patients with MIBC, including tumour tissue samples (16 FFPE TUR and 1 cystectomy; TUR tissue from patient 15 was unavailable), 17 white blood cell samples (buffy coat, BUF) and 248 body fluid samples (86 plasma, 78 UCP and 84 USN samples), spanning 86 distinct time-points in total, Figure 3.1.

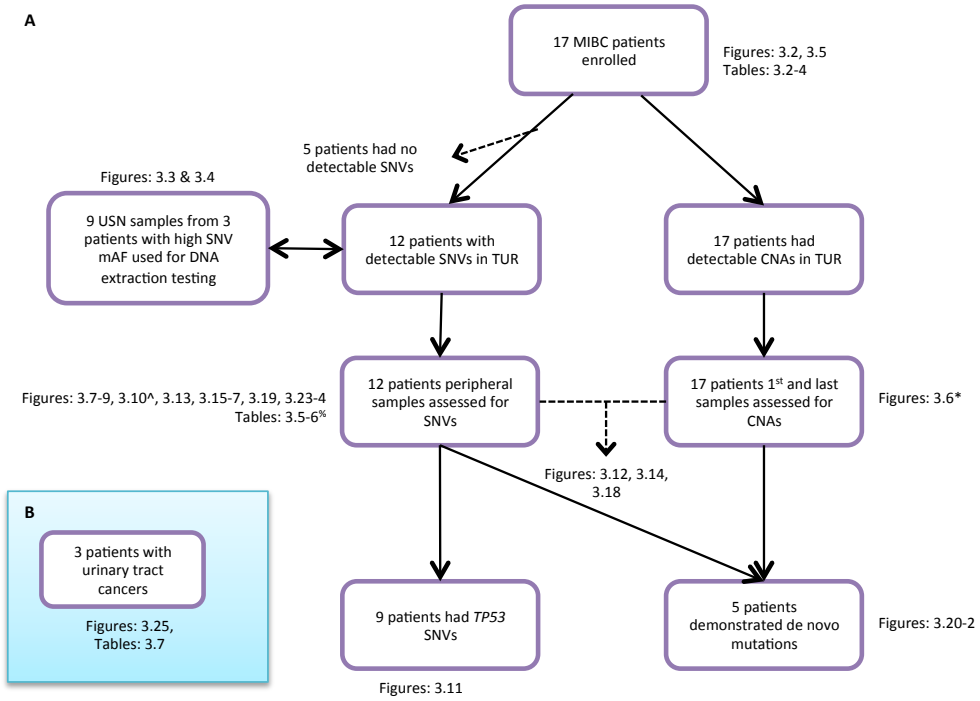



**Figure 3.1: Flow diagram of patient samples used to investigate mutDNA. A. MIBC Samples used for mutDNA analysis.** 17 patients with MIBC were recruited during the trial window and analysed according to the flow chart. *Figure 3.6 analysed samples from time-point 1 to demonstrate that CNA profiles were similar to those found in matched TUR samples. ^Figure 3.10 only time-point 2 data was used demonstrate utility as an early marker of recurrence. %Tables 3.5-6 analysed data from time-point 1. **B. Samples for investigating mutDNA kinetics in multiple urinary tract cancer disease settings.** Proof of principle analysis was conducted to demonstrate the utility of mutDNA kinetics in 3 additional urothelial carcinoma patients.

TUR samples were requested from referring hospitals whilst peripheral samples were collected at the NKI prior to administration of each cycle of NAC (Figure 3.5A). Patients were followed up for a median of 742 days (487-952 days) following initiation of NAC and 588 days (463-851 days) following definitive therapy. Details of the patient demographics, tumour characteristics and treatment are outlined in Table 3.2. DNA extraction failed in 2 samples (1 plasma and 1 USN). DNA concentration was measured by a dPCR assay

targeting the *RPP30* gene using a 97bp amplicon (Dawson *et al.*). Excluding the two failed samples, we obtained a median of 5,296 amplifiable copies/ml (ranging from 101 to 937,600 GE copies/ml), with the highest extraction yields from UCP, USN, then plasma samples (respective medians in GE copies/ml; 61,610, 5,870 and 3,550, Figure 3.2).

| Pt IDs | Age at TUR | Sex | No. of Samples Collected | | | | | TUR Grade | TUR & Imaging Stage | SNVS in TUR | Post TUR Cystoscopy | Days from TUR to NAC | Max AF at the first time-point | Definitive Treatment | Response | Final Pathology | Max AF at Last Time-point | Time to Recurrence | Max AF post NAC initiation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | TUR | BUF | PLS | UCP | USN | | | | | | | | | | | | |
| 2 | 57 | M | 1 | 1 | 4 | 4 | 4 | 3 | T3N0 | | 3 | 56 | | CR | PR | | | | |
| 7 | 48 | F | 1 | 1 | 6 | 5 | 6 | 3 | T3N0 | *TP53* | 0 | 30 | 0.000 | RC+LND | CR | T0N0 | 0.001 | | 0.001 |
| 8 | 76 | M | 1 | 1 | 5 | 4 | 5 | 3 | T3N0 | *TP53* | 0 | 44 | 0.001 | RC+LND | CR | T0N0 | 0.002 | | 0.002 |
| 9 | 71 | F | 1 | 1 | 5 | 4 | 5 | 3 | T3N0 | *PIK3CA* | 2 | 44 | 0.002 | RC+LND | SD | T3N0 | 0.003 | 378 | 0.001 |
| 11 | 49 | F | 1 | 1 | 5 | 5 | 5 | 3 | T2N0M0 | *KRAS, TP53* | 2 | 36 | 0.156 | RC+LND | CR | T0 | 0.039 | Other* | 0.167 |
| 12 | 66 | M | 1 | 1 | 4 | 4 | 4 | 3 | T2N0M0 | *PIK3CA* | 2 | 23 | 0.038 | Rad | CR | | 0.017 | 269 | 0.045 |
| 13 | 58 | M | 1 | 1 | 5 | 5 | 5 | 3 | T3N0M0 | *TP53* | 1 | 45 | 0.229 | RC+LND | SD | T2N1 | 0.018 | 507 | 0.060 |
| 15 | 66 | F | 0 | 1 | 6 | 5 | 6 | 3 | T3N1 | | 3 | 49 | 0.124 | RC+LND | PD | T3N2 | 0.427 | 293 | 0.163 |
| 18 | 56 | M | 1 | 1 | 6 | 6 | 6 | 2 | T4N0 | | 2 | 39 | | RC+LND | SD | T3N2 | | 264 | |
| 19 | 57 | M | 1 | 1 | 6 | 5 | 5 | 3 | T3N0 | *KRAS, TP53* | 1 | 51 | 0.005 | RC+LND | CR | T0N0 | 0.000 | | 0.004 |
| 21 | 64 | F | 1 | 1 | 4 | 4 | 4 | 3 | T3N0 | | 2 | 11 | | RC+LND | CR | T0 | | 466 | |
| 24 | 66 | M | 1 | 1 | 5 | 5 | 5 | 3 | T3N0 | | 1 | 32 | | RC+LND | PR | T0 N1 | | | |
| 26 | 50 | M | 1 | 1 | 6 | 6 | 6 | 3 | T3N0 | *TP53* | 1 | 35 | 0.112 | Partial RC+ Rad | SD | T3 | 0.049 | 472 | 0.211 |
| 27 | 58 | M | 1 | 1 | 4 | 3 | 4 | 3 | T3pN2 | *TP53* | 2 | 81 | 0.006 | RC | CR | T0 | 0.006 | | 0.002 |
| 29 | 59 | M | 1 | 1 | 6 | 6 | 6 | 3 | T3N0 | *TP53, FGFR3, PIK3CA* | ND | 27 | 0.003 | RC+LND | PR | T3 | 0.003 | | 0.002 |
| 32 | 65 | M | 1 | 1 | 5 | 3 | 4 | 3 | T3N0 | *KRAS, BRAF* | 1 | 83 | 0.489 | RC+LND | SD | T3 | 0.077 | 283 | 0.006 |
| 33 | 70 | M | 1 | 1 | 4 | 4 | 4 | 3 | T3N1 | *TP53, CTNNB1* | 0 | 61 | 0.002 | 1. CR 2. LND | CR | N0 | 0.001 | | 0.002 |

**Table 3.2: Demographics of 17 MIBC patients:** The median age at time of TUR was 59 with the cohort consisting of 12 males (M) and 5 females (F), in keeping with the prevalence of BC. The number of TUR, BUF, PLS, UCP and USN samples obtained for each patient are presented. Most patients with MIBC had grade 3 (G3), locally advanced disease and opted for Radical Cystectomy (RC). Furthermore, 8/17 patients had early recurrence (median 336 days, ranging from 264 to 507). One patient (*) died shortly after surgery due to surgical complication and was thus excluded from further analysis involving correlations with early recurrence outcome. TUR and imaging stage information and final radical cystectomy pathology are provided as per TNM criteria. Definitive Treatments: RC - Radical Cystectomy, CR – Chemoradiotherapy, Rad – Radiotherapy, LND – Lymph Node Dissection, ND – Not done. Cystoscopy findings: 0 - no disease present, 1 - equivocal findings, 2 - small tumour (<3cm), 3 - large tumour (>3cm), ND - not done. Pathological response categories: pCR- pathological complete response, pPR - pathological partial response, pSD - pathological stable disease, pPD - pathological progressive disease. Other*: patient died from post operative complications.

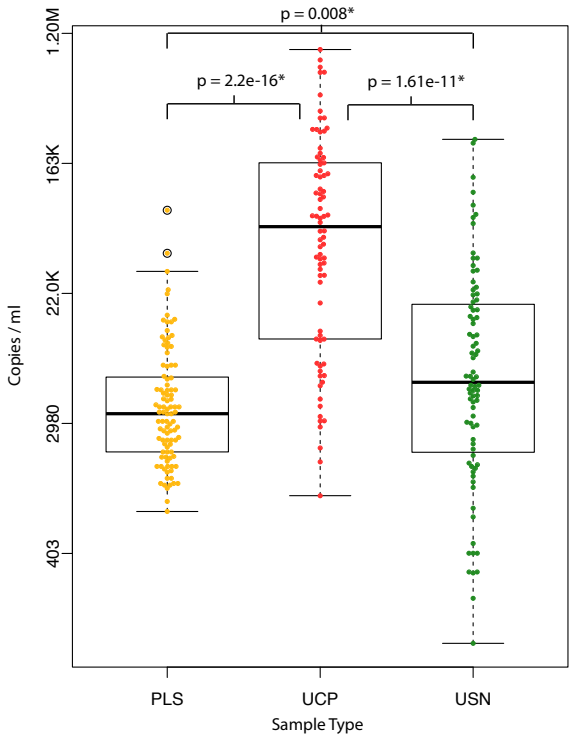



**Figure 3.2: Amount of DNA (genome equivalent copies/ml) extracted from peripheral samples.** UCP samples had a statistically higher number of DNA copies/ml of fluid ($p<0.0001$) as determined by Kruskal Wallis testing. There was also a statistically significant difference between extracted DNA copies/ml from USN and PLS. The respective medians and ranges in GE copies/ml for UCP, USN and PLS were; 61,610 (981–937,600), 5,870 (101–235,500) and 3,550 (769-79,230).

### 3.6.2 cfDNA yield and mutDNA levels are comparable when extracting with QIAmp Circulating Nucleic Acid kit and QIAsymphony (Qiagen)

During the course of the study, automated DNA extraction from urine samples, with the QIAsymphony became available. In brief, the Qiagen QIAsymphony extraction method can accommodate up to 96 samples in a single extraction run, though this is done in 4 stages of 24 samples. Instead of the silica gel membrane used in the Circulating Nucleic Acid kit, the QIAsymphony couples silica extraction with magnetic beads to extract DNA (QIagen, 2017). The automated nature of sample extraction could allow greater reproducibility, with less room for human error together with simplicity when extracting a large number of samples. Indeed, using whole blood, Laus *et al.* have demonstrated that QIAsymphony extractions have greater inter-run and intra-run reproducibility when compared with QIAmp manual extractions (Laus *et al.*, 2011). When estimating Epstein-Barr Viral loads in the blood in paired samples, they found similar viral copies/ml from both extraction methods.

To investigate the utility of QIAsymphony urinary DNA extraction, USN samples from MIBC patients in the first cohort of analysis underwent manual Circulating nucleic acid kit extraction of a 3ml aliquot and TAm-Seq. Nine samples from 3 patients (patient 1, 11 and 12) were selected based on their high AFs (>1%). A further 3ml USN aliquot of these samples was sent for automated QIAsymphony extraction through Qiagen's commercially available sample prep service (c/o Dr Anke Singer, Qiagen Sample prep service, QIAGEN GmbH, Hilden Commercial Register Düsseldorf, HRB 45822). Digital PCR using 95bp probes was performed in duplicate to quantify cfDNA. Mean dPCR *RPP30* counts were utilised for conversion to amplifiable GE copies/ml as described in Chapter 2, section 2.4.2.

Figure 3.3 shows the amplifiable GE copies/ml for paired aliquots extracted by Circulating nucleic acid kit and by QIAsymphony. To analyse concordance, a linear regression model was applied using the 'lm' function in R (R Core Team 2015). In brief, linear regression models can be used to estimate whether there is a relationship between independent variables, in this case, the method of DNA extraction. The resulting linear model has a slope of 0.778 and a y-intercept of 4059.6 GE copies/ml. The slope indicates that QIAsymphony extraction may result in higher DNA yield from paired samples, though the intercept indicates a different baseline. Taken in combination the slope predicts that with lower cfDNA levels, Circulating Nucleic Acid kits might be preferred, whilst with higher cfDNA levels (i.e. >~18,000 GE copies/ml), QIAsymphony extraction may be preferred.

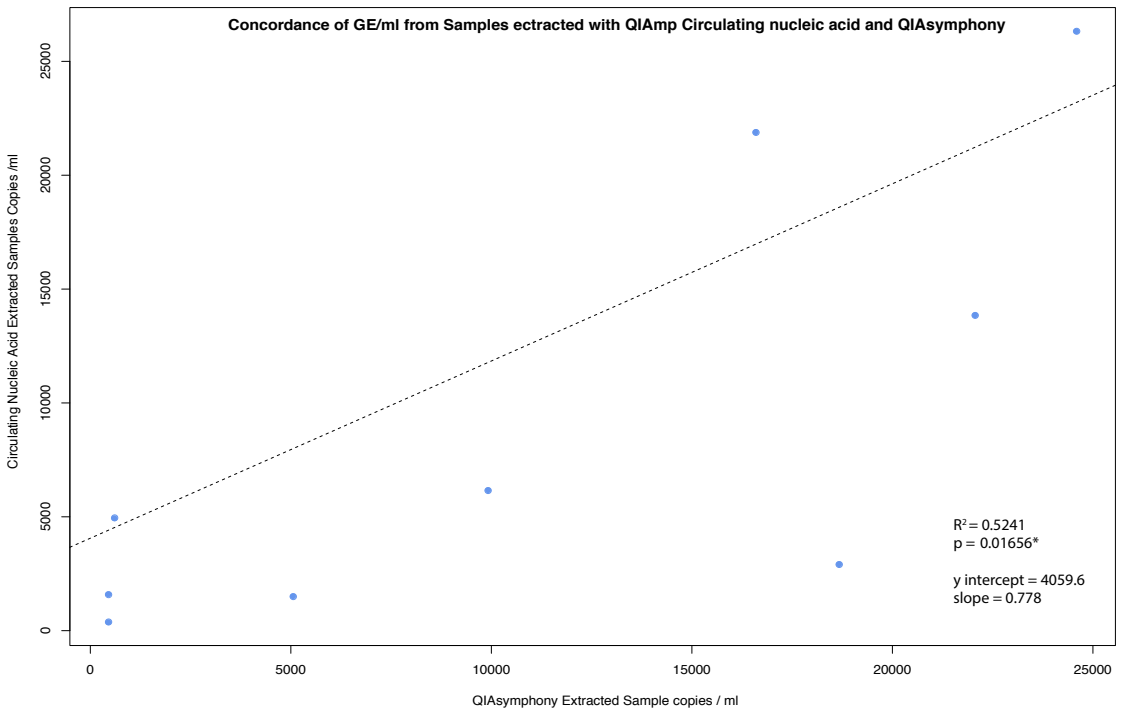



**Figure 3.3: Scatter plot of total amplifiable copies detected from replicate aliquots extracted by QIAsymphony and Circulating Nucleic Acid kits.** dPCR demonstrated no significant difference in amplifiable copies per ml of USN. The linear model is represented by a grey dotted line (slope = 0.778, adjusted $R^2$= 0.5241).

To assess the strength of any relationship between DNA yield and the extraction methods, coefficients of determination were calculated with the 'lm' function to illustrate the variance between the linear regression model and observed data. The adjusted $R^2$ was 0.5241, indicating that just 52% of the data points can be explained by the linear model described above. However, the associated p value of 0.0166 rules out a statistically significant difference in the extraction efficiencies of the two methods on the resulting DNA yield.

To investigate whether QIAsymphony DNA extraction had an effect on the mutDNA AF's, TAm-Seq was performed on DNA extracted from the paired QIAsymphony and Circulating nucleic acid kit (Qiagen) extractions as previously described by Forshew *et al.* (Forshew *et al.*) and in Chapter 2, section 2.4.5. The resulting AF's were plotted in Figure 3.4 and colour coded by SNV using R. A linear regression model was applied to determine whether there was concordance between the data and resulted in a slope of 0.847 with an intercept of 0.002. The slope suggests that higher levels of mutDNA might be obtained from the same sample by using QIAsymphony extraction instead of the circulating nucleic acid DNA extraction kits. The adjusted $R^2$ value is 0.8942 suggesting that the model can explain almost 90% of the data and that variance from the model is low. A p value of 6.358 x $e^{-12}$ confirms that there is no effect of extraction method on the AF of a sample. That

there is no significant difference between cfDNA yields nor mutDNA AF's between QIAmp Circulating nucleic acid manual extraction and automated QIAsymphony extraction corroborates the findings of Laus *et al.* (Laus *et al.*, 2011) and therefore, QIAsymphony was utilised for further cfDNA extractions.

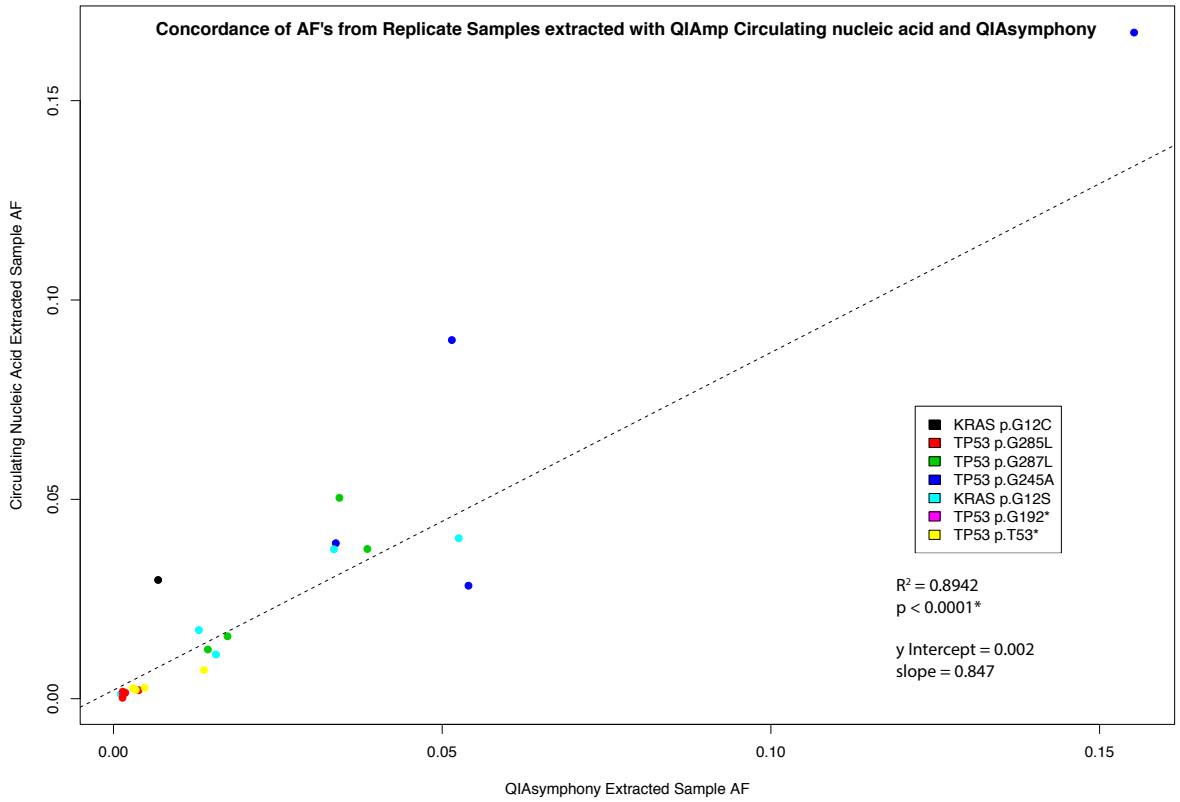



**Figure 3.4: Scatter plot of mAFs detected from replicate aliquots extracted by QIAsymphony and Circulating Nucleic Acid kits.** TAm-Seq was used to estimate the mutDNA load for each sample. The line model is represented by a grey dotted line (slope = 0.847, adjusted $R^2$ = 0.8942).

### 3.6.3 Detection of DNA alterations in TUR samples from MIBC patients undergoing NAC

For analysis of single nucleotide variant mutations (SNVs), the previously described bladder-specific sequencing panel was used. When applied to MIBC, we expect the panel to detect SNVs in approximately 72% of cases based on the frequency of SNV mutations covered in published data (Table 3.1). TUR sample DNA libraries were submitted for 125bp paired end NGS using 1 lane of a HiSeq 2500/4000 (Illumina) instrument. The HiSeq run resulted in approximately 257.2 million reads. TUR samples had a median sequencing depth of 9600X.

A total of 22 SNVs were detected across 12 of the 16 patients (75%) Table 3.2 and Figure 3.5B. The most frequent mutations were detected in *TP53* (14 mutations across 10 patients), followed by *KRAS* (3 mutations), then *PIK3CA* (2 mutations). One SNV each was also detected in the *BRAF*, *CTNNB1* and *FGFR3* genes. These findings agree with previous studies annotating the frequency of SNVs in MIBC (Cancer Genome Atlas Research, 2014; Forbes *et al.*) and the predicted mutation rate based on our panel design. Though the numbers of SNVs detected in TUR samples varied amongst patients, there was no correlation with SNVs in TUR per patient and clinical outcome.

For CNAs analysis in TUR, the mean number of sequencing reads achieved for each sample was 14.2 million. Gross genome wide CNAs were detected in all tumour specimens, including for those 5 patients in whom no SNVs were observed (Figure 3.5B). Across the 16 TUR samples, we detected focal *CDKN2A loss (37.5%), E2F3/SOX4 gain (37.5%), PPARG gain (25.0%), YWHAZ gain (18.8%), CREBBP loss (12.5%), MYCL1 gain (12.5%)* and CNAs of other BC genes, as previously shown (Cancer Genome Atlas Research, 2014).

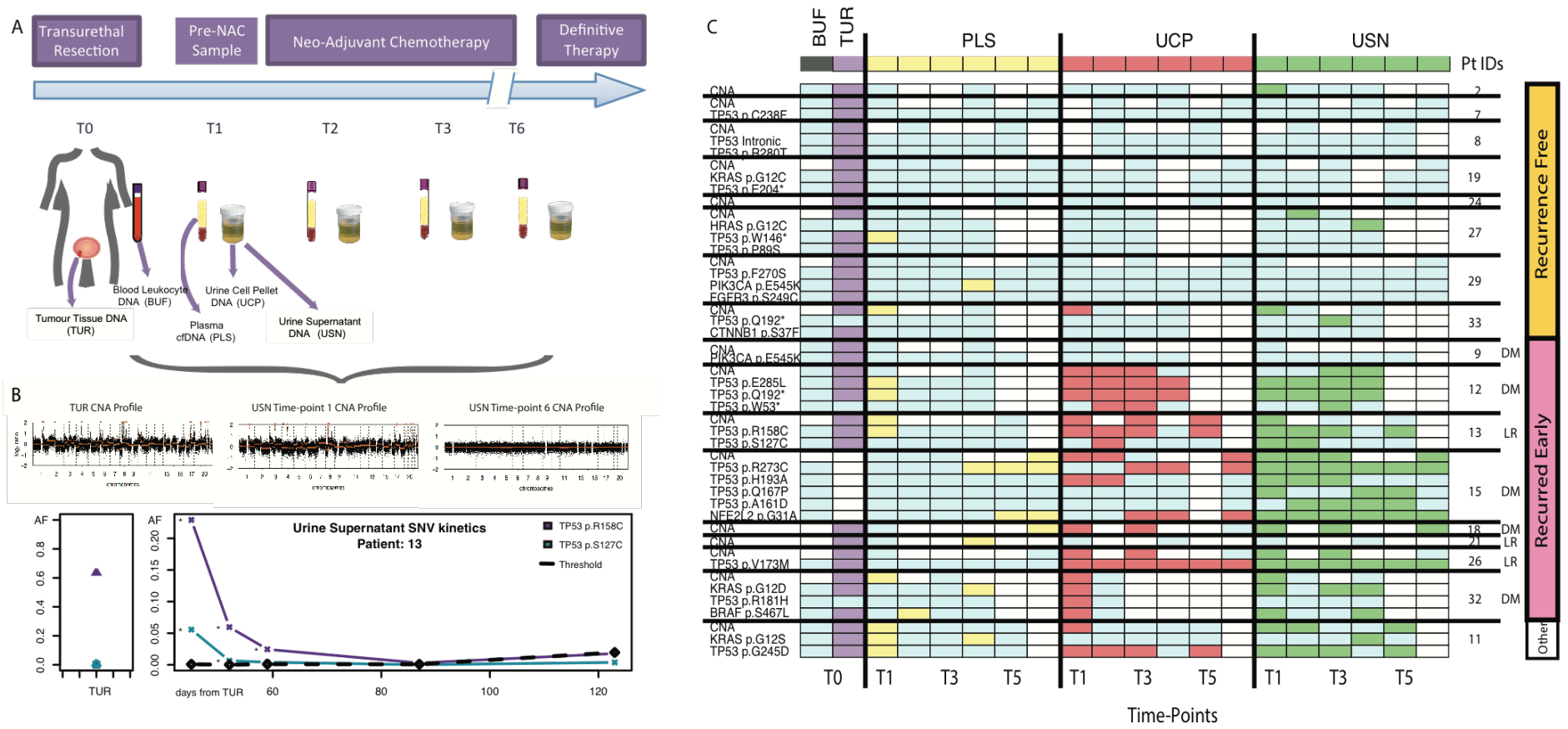



**Figure 3.5: Analysis of Longitudinal mutDNA kinetics in MIBC. A. Study design.** 17 patients were enrolled at the NKI for mutDNA analysis whilst undergoing NAC. TUR was performed at the NKI or at regional institutions. PLS, USN and UCP were taken on one occasion before the initiation of NAC and subsequently on each chemotherapy visit, prior to definitive therapy. Examples of longitudinal SNV and CNA data are shown. **B. Grid depicting mutDNA detection across all patients and time-points.** The right y-axis shows patients grouped by recurrence status with recurrence type adjacent; DM - distant metastasis, LR - Local recurrence. Individual mutations are listed on the left y axis. The x-axis shows sample-types (top) and time-points (bottom). Each cell of the grid indicates the result of mutation analysis. White cells correspond to unavailable samples and light blue cells to samples where analysis did not detect mutDNA. Purple, yellow, red and green cells correspond to TUR, PLS, UCP and USN samples (respectively) where mutDNA was detected. No mutations were detected in BUF (Grey). Raw AFs for the grid are provided in **Table 3.4**.

| Pt ID | Reaction Type | Recurrance status | T0.BUF | T0.TUR | T1.PLS | T2.PLS | T3.PLS | T4.PLS | T5.PLS | T6.PLS | T1.UCP | T2.UCP | T3.UCP | T4.UCP | T5.UCP | T6.UCP | T1.USN | T2.USN | T3.USN | T4.USN | T5.USN | T6.USN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2 | CNA | N | 25330 | 8606 | 91 | | | 513 | | | 63940 | 3030 | 37500 | 7473 | | | 13340 | 1372 | 3030 | 3030 | | |
| 2 | SNV | N | 7576 | 7576 | 90 | 303 | 192 | 256 | | | 31950 | 18660 | 37450 | 7576 | | | 13350 | 909 | 1292 | 1077 | | |
| 7 | CNA | N | 14000 | 112700 | 359 | | | 1030 | | 308 | | | | 13720 | | 8461 | 3030 | 3030 | | 3023 | | 3030 |
| 7 | SNV | N | 7576 | 28170 | 179 | 303 | 1585 | 303 | 303 | 154 | | 7576 | 10960 | 7576 | 7703 | 7576 | 1028 | 2277 | 4249 | 928 | 909 | 1454 |
| 8 | CNA | N | 14970 | 143000 | 949 | 3030 | | | 0 | | | 14250 | | | 21960 | | 179 | 3030 | | | 3030 | |
| 8 | SNV | N | 7576 | 35710 | 303 | 467 | 438 | 618 | 0 | | | 7576 | 3805 | 7576 | 10980 | | 90 | 909 | 2464 | 462 | 909 | |
| 19 | CNA | N | | 15150 | 513 | | | | | 256 | 5886 | | | | | 52940 | 3030 | | | | | 2154 |
| 19 | SNV | N | 10080 | 15150 | 321 | 321 | 1115 | 1936 | 1641 | 167 | 7576 | 1489 | 7576 | | 1237 | 26460 | 1705 | 1782 | 4769 | | 282 | 1372 |
| 24 | CNA | N | | 4788 | 205 | | | | 513 | | 10800 | | | | 0 | | 3030 | | | | 51 | |
| 24 | SNV | N | 7333 | 7576 | 141 | 333 | 949 | 256 | 167 | 192 | 7576 | 7576 | 7576 | 1124 | 777 | | 2462 | 1179 | 9731 | 64 | 64 | |
| 27 | CNA | N | | 7576 | 974 | | 1333 | | | | 0 | | 29140 | | | | 2564 | | 11930 | | | |
| 27 | SNV | N | 9939 | 7576 | 449 | 667 | 641 | 1295 | 141 | | 1442 | 7576 | 14580 | | | | 1179 | 2538 | 59520 | 33330 | | |
| 29 | CNA | N | | 38640 | 462 | | | | | 308 | 11110 | | | | | 0 | 3030 | | | | | 3030 |
| 29 | SNV | N | 10030 | 38610 | 192 | 154 | 333 | 474 | 962 | 154 | 7576 | 7576 | 7576 | 7576 | 1244 | 662 | 9051 | 5513 | 2064 | 1205 | 2808 | 2141 |
| 33 | CNA | N | | 45620 | 513 | | | 615 | | | 21390 | | | 0 | | | 1179 | | | 1436 | | |
| 33 | SNV | N | 15550 | 45660 | 333 | 333 | 295 | 385 | 256 | | 10700 | 7576 | 2306 | 565 | | | 654 | 3051 | 1269 | 782 | | |
| 9 | CNA | Y | 16300 | 99390 | 436 | | | 333 | | | 26150 | | | 3030 | | | 3030 | | | 3030 | | |
| 9 | SNV | Y | 7576 | 24880 | 218 | 269 | 231 | 167 | 282 | | 13070 | 7576 | 7576 | 5797 | | | 897 | 603 | 654 | 141 | 397 | |
| 12 | CNA | Y | 19150 | 321200 | 564 | | | 273 | | | 6500 | 3209 | 3024 | 5600 | | | 3030 | 3466 | 9767 | 1049 | | |
| 12 | SNV | Y | 7576 | 80000 | 282 | 303 | 303 | 141 | | | 7576 | 7576 | 7486 | 7576 | | | 1579 | 1313 | 909 | 909 | | |
| 13 | CNA | Y | | 24760 | 462 | | | | 462 | | 4545 | | | | 0 | | 2667 | | | | 103 | |
| 13 | SNV | Y | 7879 | 24750 | 192 | 923 | 1103 | 359 | 205 | 64 | 2185 | 7606 | 1066 | 2182 | 2136 | | 1231 | 7718 | 962 | 1128 | 51 | |
| 15 | CNA | Y | | | 154 | 513 | 410 | | | 462 | 5779 | 3030 | 22410 | | | 16450 | 3030 | 3267 | 3030 | | | 3764 |
| 15 | SNV | Y | 16560 | | 282 | 205 | 244 | 192 | 282 | 192 | 7576 | 867 | 11210 | 28650 | | 8227 | 4359 | 16810 | 7846 | 5308 | 25770 | 18480 |
| 18 | CNA | Y | | 16650 | 308 | | | | | 308 | 0 | | | | | 0 | 718 | | | | | 615 |
| 18 | SNV | Y | 8348 | 16640 | 154 | 282 | 449 | 321 | 423 | 205 | 409 | 621 | 1333 | 196 | 5823 | 618 | 244 | 103 | 2744 | 936 | 418 | 256 |
| 21 | CNA | Y | | 7424 | 256 | | | 769 | | | 21130 | | | 52700 | | | 3030 | | | 2103 | | |
| 21 | SNV | Y | 10100 | 7424 | 167 | 295 | 141 | 385 | 103 | | 10560 | 7576 | 6385 | 26380 | | | 3615 | 244 | 231 | 897 | | |
| 26 | CNA | Y | | 73940 | 205 | | | | | 1282 | 4545 | | | | | 4545 | 3030 | | | | | 821 |
| 26 | SNV | Y | 9894 | 74070 | 103 | 77 | 385 | 372 | 77 | 564 | 2467 | 1505 | 7576 | 7033 | 7576 | 6877 | 2128 | 141 | 885 | 115 | 1051 | 513 |
| 32 | CNA | Y | | 23200 | 256 | | 974 | | 462 | 1436 | 4545 | | | | | | 923 | | | 154 | | |
| 32 | SNV | Y | 11090 | 23200 | 128 | 218 | 449 | 205 | 218 | 808 | 7576 | 5234 | | | | | 641 | 22470 | 1782 | 51 | | |
| 11 | CNA | Other | 27460 | 466700 | 256 | 394 | 1182 | 744 | 256 | | 3333 | 13140 | 3615 | 5764 | 4591 | | 3333 | 3333 | 3033 | 3029 | 5115 | |
| 11 | SNV | Other | 7576 | 116700 | 128 | 303 | 303 | 303 | 128 | | 13110 | 7576 | 7576 | 7576 | 7576 | | 872 | 449 | 474 | 909 | 909 | |

**Table 3.3: Grid depicting total GE copies inputted per reaction across all patients and time-points.** The format of the grid matched that of **Figure 3.5**.

| Mut ID | Pt ID | Recurrance status | T0.BUF | T0.TUR | T1.PLS | T2.PLS | T3.PLS | T4.PLS | T5.PLS | T6.PLS | T1.UCP | T2.UCP | T3.UCP | T4.UCP | T5.UCP | T6.UCP | T1.USN | T2.USN | T3.USN | T4.USN | T5.USN | T6.USN |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CNA | 2 | N | ND | 1 | ND | | | ND | | | ND | ND | ND | ND | | | 1 | ND | ND | ND | | |
| CNA | 7 | N | ND | 1 | ND | | | ND | | ND | | | | ND | | ND | ND | ND | | ND | | ND |
| TP53 p.C238F | 7 | N | ND | 0.2327 | ND | ND | ND | ND | ND | ND | | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| CNA | 8 | N | ND | 1 | | ND | | | ND | | | ND | | | ND | | | ND | | | ND | |
| TP53 Intronic | 8 | N | ND | 0.0086 | ND | ND | ND | ND | ND | | | ND | ND | ND | ND | | ND | ND | ND | ND | ND | |
| TP53 p.R280T | 8 | N | ND | 0.7903 | ND | ND | ND | ND | ND | | | ND | ND | ND | ND | | ND | ND | ND | ND | ND | |
| CNA | 19 | N | | 1 | ND | | | | | ND | ND | | | | | ND | ND | | | | | ND |
| KRAS p.G12C | 19 | N | ND | 0.5853 | ND | ND | ND | ND | ND | ND | ND | ND | ND | | ND | ND | ND | ND | ND | | ND | ND |
| TP53 p.E204* | 19 | N | ND | 0.0204 | ND | ND | ND | ND | ND | ND | ND | ND | ND | | ND | ND | ND | ND | ND | | ND | ND |
| CNA | 24 | N | | 1 | ND | | | | ND | | ND | | | | ND | | ND | | | | ND | |
| CNA | 27 | N | | 1 | ND | | ND | | | | ND | | ND | | | | ND | | ND | | | |
| HRAS p.G12C | 27 | N | ND | ND | ND | ND | ND | ND | ND | | ND | ND | ND | | | | ND | ND | ND | 0.0062 | | |
| TP53 p.W146* | 27 | N | ND | 0.8319 | 0.0061 | ND | ND | ND | ND | | ND | ND | ND | | | | ND | ND | ND | ND | | |
| TP53 p.P89S | 27 | N | ND | 0.0162 | ND | ND | ND | ND | ND | | ND | ND | ND | | | | ND | ND | ND | ND | | |
| CNA | 29 | N | | 1 | ND | | | | | ND | ND | | | | | ND | ND | | | | | ND |
| TP53 p.F270S | 29 | N | ND | 0.3324 | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| PIK3CA p.E54KL | 29 | N | ND | 0.3827 | ND | ND | ND | 0.0053 | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| FGFR3 p.S249C | 29 | N | ND | 0.2312 | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND |
| CNA | 33 | N | | 1 | 1 | | | ND | | | 1 | | | ND | | | 1 | | | ND | | |
| TP53 p.Q192* | 33 | N | ND | ND | ND | ND | ND | ND | ND | | ND | ND | ND | ND | | | ND | ND | 0.0088 | ND | | |
| CTNNB1 p.S37F | 33 | N | ND | 0.0428 | ND | ND | ND | ND | ND | | ND | ND | ND | ND | | | ND | ND | ND | ND | | |
| CNA | 9 | Y | ND | 1 | ND | | | ND | | | ND | | | | | | ND | | | | | |
| PIK3CA p.E545K | 9 | Y | ND | 0.0072 | ND | ND | ND | ND | ND | | ND | ND | ND | ND | | | ND | ND | ND | ND | ND | |
| CNA | 12 | Y | ND | 1 | ND | | | ND | | | 1 | 1 | 1 | ND | | | ND | ND | 1 | 1 | | |
| TP53 p.E285L | 12 | Y | ND | 0.4272 | 0.0120 | ND | ND | ND | | | 0.0216 | 0.0450 | 0.0729 | 0.0114 | | | 0.0375 | 0.0111 | 0.0403 | 0.0172 | | |
| TP53 p.Q192* | 12 | Y | ND | 0.4784 | 0.0152 | ND | ND | ND | | | 0.0218 | 0.0435 | 0.0826 | 0.0113 | | | 0.0376 | 0.0157 | 0.0504 | 0.0124 | | |
| TP53 p.W53* | 12 | Y | ND | ND | ND | ND | ND | ND | | | ND | 0.0099 | 0.0355 | ND | | | ND | ND | 0.007 | ND | | |
| CNA | 13 | Y | | 1 | 1 | | | | ND | | 1 | | | | 1 | | 1 | | | | ND | |
| TP53 p.R158C | 13 | Y | ND | 0.6370 | 0.0071 | ND | ND | ND | ND | | 0.0136 | 0.0438 | 0.0241 | ND | 0.0185 | | 0.2291 | 0.0596 | 0.0245 | ND | 0.0179 | |
| TP53 p.S127C | 13 | Y | ND | 0.0093 | ND | ND | ND | ND | ND | | ND | 0.0090 | ND | ND | ND | | 0.0557 | 0.0065 | ND | ND | ND | |
| CNA | 15 | Y | | 1 | ND | ND | ND | | | 1 | 1 | 1 | ND | | | 1 | 1 | 1 | | | | 1 |
| TP53 p.R273C | 15 | Y | ND | | ND | ND | ND | 0.0136 | 0.0331 | 0.0831 | ND | ND | 0.0130 | 0.0113 | | 0.0701 | 0.0102 | 0.0122 | 0.2237 | 0.3241 | 0.5570 | 0.3380 |
| TP53 p.H193A | 15 | Y | ND | | ND | ND | ND | ND | ND | ND | 0.0160 | 0.0320 | ND | ND | | ND | 0.1237 | 0.1634 | 0.0201 | ND | ND | ND |
| TP53 p.Q167P | 15 | Y | ND | | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | | ND | 0.0056 | ND | ND | 0.0053 | 0.0053 | ND |
| TP53 p.A161D | 15 | Y | ND | | ND | ND | ND | ND | ND | ND | ND | ND | ND | ND | | ND | ND | 0.0144 | 0.0188 | 0.0178 | 0.0050 | ND |
| NFE2L2 p.G31A | 15 | Y | ND | | ND | ND | ND | ND | 0.0128 | 0.0484 | ND | ND | 0.0069 | 0.0076 | | 0.0428 | 0.0110 | 0.0098 | 0.1684 | 0.2443 | 0.3963 | 0.4272 |
| CNA | 18 | Y | | 1 | ND | | | | | 1 | 1 | | | | | ND | 1 | | | | | 1 |
| CNA | 21 | Y | | 1 | ND | | | 1 | | | ND | | | ND | | | ND | | | ND | | |
| CNA | 26 | Y | | 1 | ND | | | | | ND | 1 | | | | | ND | 1 | | | | | ND |
| TP53 p.V173M | 26 | Y | ND | 0.8883 | ND | ND | ND | ND | ND | ND | 0.0665 | 0.2024 | 0.0318 | 0.0400 | 0.0135 | 0.0159 | 0.1121 | 0.2105 | 0.4843 | 0.0691 | 0.0419 | 0.0487 |
| CNA | 32 | Y | | 1 | 1 | | ND | | | | 1 | | | | | | 1 | | | ND | | |
| KRAS p.G12D | 32 | Y | ND | 0.2859 | ND | ND | ND | 0.0088 | ND | ND | 0.4892 | ND | | | | | 0.3460 | ND | 0.0071 | 0.0444 | | |
| TP53 p.R181H | 32 | Y | ND | ND | ND | ND | ND | ND | ND | ND | 0.0059 | ND | | | | | ND | ND | ND | ND | | |
| BRAF p.S467L | 32 | Y | ND | 0.0542 | ND | 0.0057 | ND | ND | ND | ND | 0.2315 | ND | | | | | 0.0546 | ND | ND | 0.0771 | | |
| CNA | 11 | Other | ND | 1 | 1 | ND | ND | ND | ND | | 1 | ND | ND | ND | ND | | 1 | 1 | ND | ND | 1 | |
| KRAS p.G12S | 11 | Other | ND | 0.2306 | 0.0407 | ND | ND | 0.0068 | ND | | ND | ND | ND | ND | ND | | ND | ND | ND | 0.0067 | ND | |
| TP53 p.G245D | 11 | Other | ND | 0.9387 | 0.1557 | ND | ND | ND | ND | | 0.0459 | 0.0207 | 0.0162 | ND | 0.0309 | | 0.0900 | 0.1671 | 0.0284 | ND | 0.0390 | |

**Table 3.4: Grid depicting mutDNA AFs across all patients and time-points.** The format of the grid matched that of **Figure 3.5**. Where tested, data for CNAs were entered manually as 1 or ND depending on whether or not they were called. ND – not detected

### 3.6.4 Comparison of genomic profiles in tumour and pre-NAC peripheral samples

Pre-NAC samples were collected from patients 1-2 months following their TUR but immediately prior to starting NAC. To detect mutDNA in peripheral fluid samples, TAM-Seq and sWGS libraries were submitted for 125bp paired end re-sequencing on 3 lanes of a Illumina HiSeq 2500. Peripheral samples had a median depth of 7600X for SNV analysis and a mean of 13.6 million reads / sample for CNA analysis. Of the 12 patients with SNVs detected in TUR, one or more identical SNVs were also detected in 30.8% (4/12) 46.2% (5/12) and 46.2% (5/12) of pre-NAC plasma, UCP and USN samples respectively. Similarly, CNAs were observed in 25% (4/16), 53.3% (8/15) and 50% (8/16), and of pre-NAC plasma, UCP and USN samples of all patients with available samples. Figure 3.6 demonstrates that many of these CNAs occurred in multiple fluid types of some of the patients and, that when a mutDNA CNA signal was detectable in peripheral fluids, the CNA profile subjectively matched the CNA profile of the corresponding TUR. Linear modeling of CNA profiles from each peripheral fluid, in which a CNA was called and the matched TUR resulted in a median adjusted $R^2$ of 0.4990 (range 0.0011 – 0.9777) and is likely to represent the reduced prominence of CNA profiles in the peripheral samples, possibly due to higher levels of germ-line DNA in urine and plasma samples, due to successful removal of the bulk of disease by TUR, or due to spatial and temporal tumour heterogeneity (discussed below). When combining both methods, mutations (SNVs and CNAs) were detected in the first (pre-NAC) time-point in 58.8% (10/17) patients. MutDNA was present at the pre-NAC time-point in 35.3% (6/17), 47.1% (8/17) and 52.9% (9/17) of plasma, UCP and USN samples respectively.

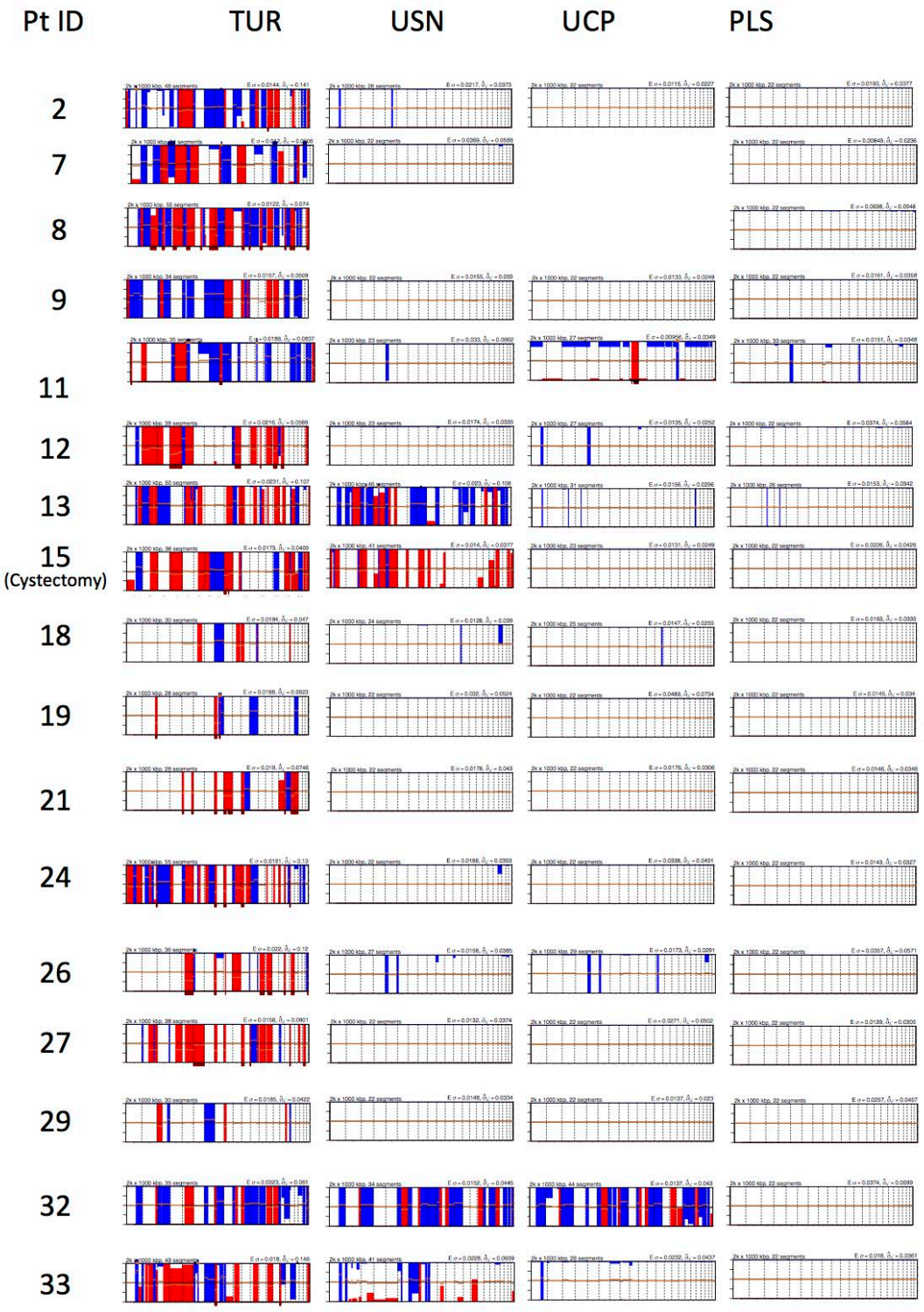



**Figure 3.6: sWGS data for TUR and peripheral fluids taken prior to NAC for 17 MIBC patients.** Genome wide log2 ratio plots depict genomic position (x axis) and log 2 ratio scores using 1Mb bins (y axis). CNAs were deemed to be present using CGHcall (van de Wiel *et al.*, 2007) in either 1Mb or 50kb (as described in section 3.4.4). All TUR samples (the cystectomy sample for pt15) contained CNAs. CNAs were detected in CNAs were detected in 25% (4/16), 53.3% (8/15) and 50% (8/16), and of pre-NAC plasma, UCP and USN samples respectively.

### 3.6.5 Presence of mutDNA in pre-NAC peripheral samples has poor correlation to outcome.

To investigate the utility of mutDNA detection as a predictive biomarker for response to NAC in patients with MIBC, mutDNA detection in pre-NAC samples was correlated with clinical outcome. Presence of mutDNA was defined as its detection in one or more sample type(s) at an AF, greater than 1/GE copies inputted per reaction (our input threshold, raw data in Table 3.3) and higher than 0.5% (our technical threshold). This technical threshold was previously described at 2% for the identification, and 0.14% for the detection of SNVs by Forshew *et al.* in 2012 (Forshew *et al.*, 2012). Whilst, AUC curves (Figure 3.7) show that a number of technical thresholds could have been used, we utilised a working threshold of 0.5% for our proof of principle study as this ensured 100% specificity, and meant that in buffy coat samples AFs at all genomic co-ordinates with non-reference calls were below our threshold, Figure 3.8.

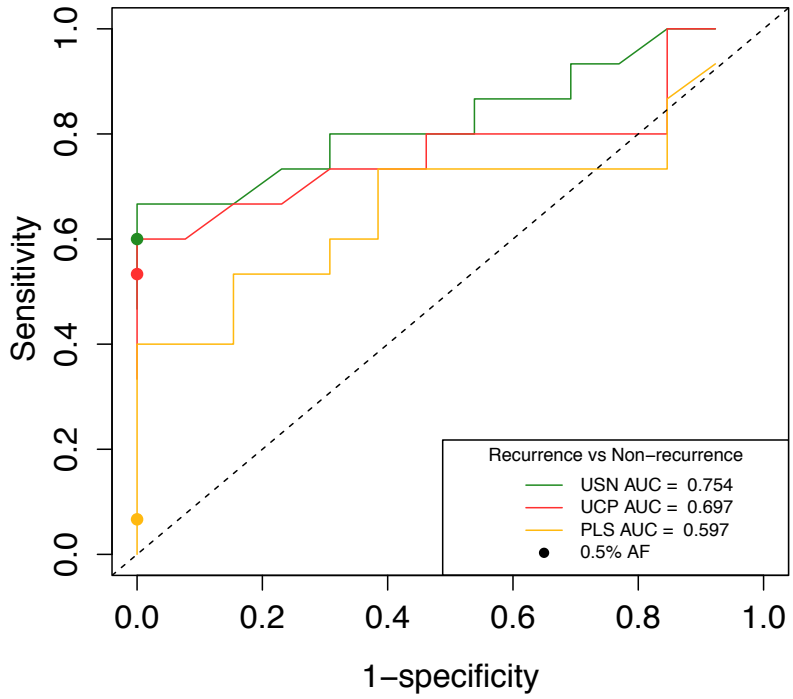



**Figure 3.7: AUC plots showing the trade off between sensitivity and specificity for mutDNA to predict recurrence at various threshold cut-offs.** Prediction of recurrence was assessed using data from samples obtained immediately prior to the 2nd cycle of NAC. The ability of mutDNA to predict recurrence is detailed in section 3.6.6. Circles represent values obtained when using our threshold of 0.5% AF for mutation calling.

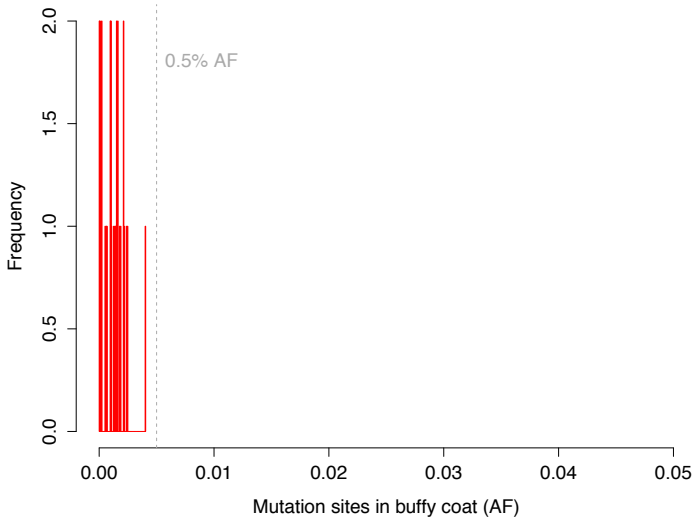



**Figure 3.8: Buffy coat AFs at TAm-Seq called locations with technical threshold plotted as 0.5%.**

According to these criteria, MutDNA was detected in pre-NAC samples in 6 patients who showed complete or partial pathological response and in 4 patients who showed no pathological response or progression according to final pathology at radical cystectomy. 5 patients who responded and 2 patients who did not, showed no evidence of mutDNA at this time-point. The sensitivity and specificity of mutDNA detection in pre-NAC samples to predict response to chemotherapy was 66.6% and 45.5% respectively, with positive and negative predictive values of 40% and 71.4% respectively. Therefore, the detection of mutDNA in pre-NAC samples did not correlate with early response to NAC in this sample set (Table 3.2 and Table 3.5).

| | | **Pathological Response to NAC** | |
|---|---|---|---|
| | | pCR/pPR | pSD/pPD |
| mutDNA status | Detected | 6 | 4 |
| | Not Detected | 5 | 2 |

**Table 3.5: Sensitivity and specificity for mutDNA detection to predict response to NAC.** Detection of mutDNA in pre-NAC samples did not correlate with response to NAC. Response categories: pCR – Patholigcal complete response, pPR – Pathological partial response, pSD – Pathological stable disease, pPD – Pathological progressive disease. Overall sensitivity and specificity were 54.5% (95% CI: 23% - 83%) and 33.3% (95% CI: 4% - 78%) with positive predict value and negative predictive values of 60% and 29% respectively.

However, the presence of mutDNA pre-NAC seemed to be more in keeping with the presence of residual disease as defined by flexible cystoscopy. Indeed, 16/17 patients had diagnostic flexible cystoscopy following TUR, before NAC was commenced. On cystoscopy, 8/16 patients had obvious residual tumour present, and mutDNA was detected in the peripheral samples in 6/8 patients with residual tumour and in 1/3 patients with no obvious residual tumour. mutDNA was detected in a further 3/5 patients with equivocal cystoscopy findings, Table 3.6. The sensitivity and specificity of mutDNA detection in pre-NAC samples to predict residual disease detection by flexible cystoscopy was 75% and 66.6% respectively, with positive predictive value and negative predictive values of 85.7% and 50% respectively.

| | | **Flexible Cystoscopy Findings** | |
|---|---|---|---|
| | | Residual tumour | NAD |
| mutDNA status | Detected | 6 | 1 |
| | Not Detected | 2 | 2 |

**Table 3.6: Sensitivity and specificity for mutDNA detection in pre-NAC samples.** Overall sensitivity and specificity were 75% (95% CI: 35% - 97%) and 66.6% (95% CI: 9% - 99%) with positive predict value and negative predictive values of 85.7% and 50% respectively.

### 3.6.6 Presence of mutDNA during NAC is associated with recurrence

The majority of mutDNA was detected in samples taken from patients who recurred after definitive therapy (Figure 3.5B). 8 patients in our cohort recurred, with a median time to recurrence of 336 days (maximum 507 days) from initial TUR. One patient (patient 11), died due to a post-operative complication of radical cystectomy and was categorised as “other” and excluded from further analysis. Patients 8 and 29 developed new primary malignant melanoma and lung adenocarcinoma tumours at 882 and 175 days respectively during follow up. These patients were censored at the date of new tumour diagnosis to be recurrence free for the purposes of our analysis. All other patients were censored at the time of analysis. Therefore, a total of 8 patients were recurrence free after a median follow up of 781 days after TUR (maximum 1008 days). Overall, 90 SNVs were detected in 219 mutant-time-point analyses, and 26 CNAs were detected in 54 samples, from the 8 patients who recurred (Figure 3.5B). However, only 4 SNVs were detected in 193 mutant-time-point analyses and 4 CNAs were detected in 55 samples tested from the 8 non-recurring patients. Chi-Squared comparison at each time-point showed a significantly

greater mutDNA detection rate in patients that recurred as compared to those that did not at time-points 1, 2, 3 and 5, (p= 0.0058 at T1, p= 0.0055 at T2, p= 0.0130 at T3 and p= 0.0272 at T5 after Bonferroni correction for multiple testing). Also, SNV AFs at each time-point were significantly higher in patients who recurred compared to those who did not (Kolmogorov-Smirnov, p<0.022 at all time-points after Bonferroni correction, see Figure 3.9).

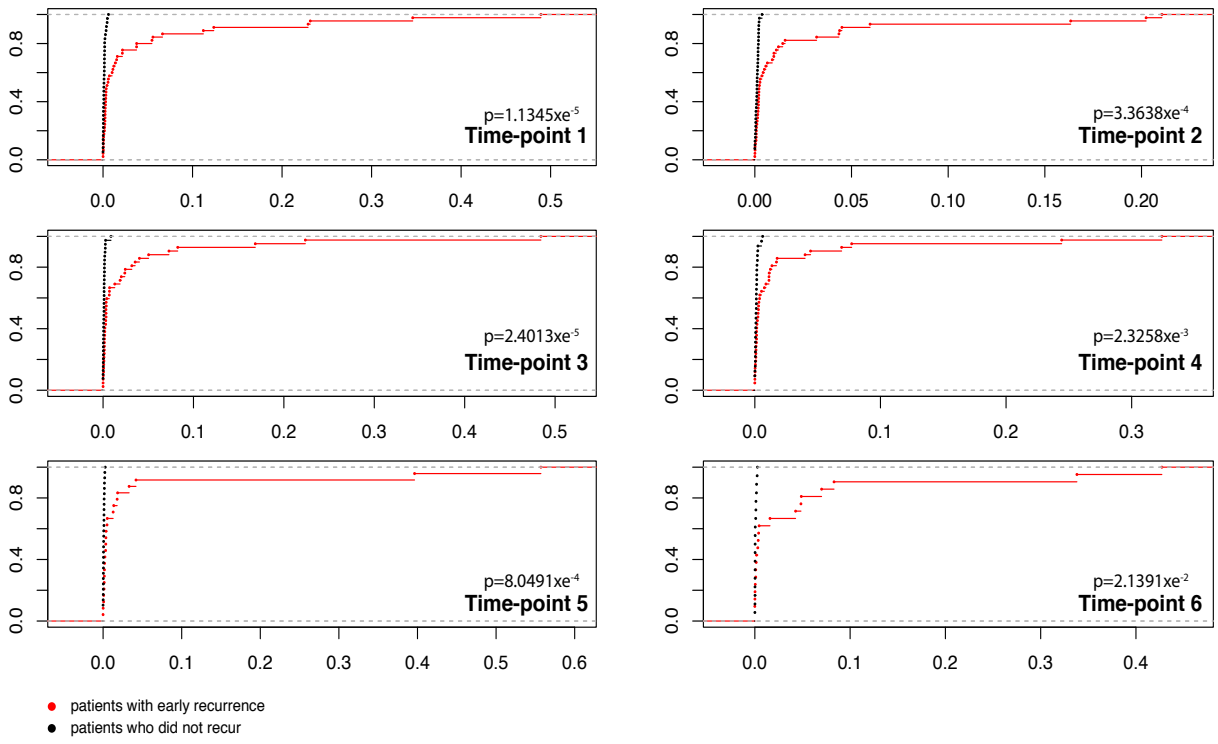



**Figure 3.9: Empirical Cumulative Distribution Function plot of SNV AFs for all samples at each time-point from patients who recurred (red) and those who did not (black).** Each time-point was analysed independently. Kolmogorov-Smirnov testing was used to compare SNV AFs between recurrence and non-recurrence groups, with Bonferroni correction for multiple testing.

To investigate potential utility of mutDNA analysis for the prediction of recurrence in patients with MIBC, we analysed the persistence of mutDNA, in peripheral samples taken at an early on-treatment time-point (i.e. just prior to the administration of the 2$^{nd}$ NAC cycle). We utilised TAm-Seq to detect mutDNA at this time-point (CNA data for samples taken at 2$^{nd}$ NAC cycle was not available). As such 5 patients were precluded from further analysis as SNVs were not detected in their tumour nor peripheral samples. Presence of mutDNA was defined as its detection in one or more sample type(s) at an AF, greater than our input threshold (1/GE copies inputted per reaction, Table 3.3) and higher than our technical threshold (0.5% as described in section 3.6.5). According to these criteria, mutDNA was present at the 2$^{nd}$ NAC cycle visit in 5 of the 6 patients that recurred,

whereas it was not detected in any of the cases that did not recur, 83% sensitivity (95% CI: 36-100%) and 100% specificity (95% CI: 42-100%), Figure 3.10A and B). All of the patients with detected mutDNA at the 2nd NAC cycle had disease recurrence, with a median time to recurrence of 293 days (Figure 3.10), whereas patients in whom mutDNA was not detected had a low recurrence rate (p=0.006 using log rank test Figure 3.10 resulting in 100% positive predictive value and 85.7% negative predictive value, Figure 3.10B).

For the single patient (patient 9) who recurred despite not having detectable mutDNA at this time-point in any peripheral sample, the tumour had a *PIK3CA* E545K mutation that was present at an AF of only 0.7%. It is likely that this mutation represents a minor subclone of cells in the tumour and therefore may not be present in recurrent tumour. As a biomarker, mutDNA detection in samples taken at cycle 2 of NAC offered a median lead-time over radiological detection of recurrence of 243 days (range 182-455 days) in our data. This association was primarily driven by detection of *TP53* SNVs in the urinary samples, where 4/5 patients that recurred had mutDNA while only 1/5 patients had a BRAF SNV detected in their plasma (Figure 3.10C, raw data in Table 3.4). In a subgroup analysis of the 9 patients with *TP53* SNVs detected in tumour specimens, all 4 patients who had *TP53* SNVs detected in their USN sample at the 2nd cycle of NAC recurred, with a median time to recurrence of 271 days. The 5 patients who did not have a *TP53* SNV detected in their USN at the 2nd cycle of NAC did not recur. MutDNA detection in this subgroup would have a sensitivity and specificity of 100% and a positive predictive value (95% CI: 28-100%) and negative predictive value of 100% (95% CI: 36-100%), Figure 3.11.

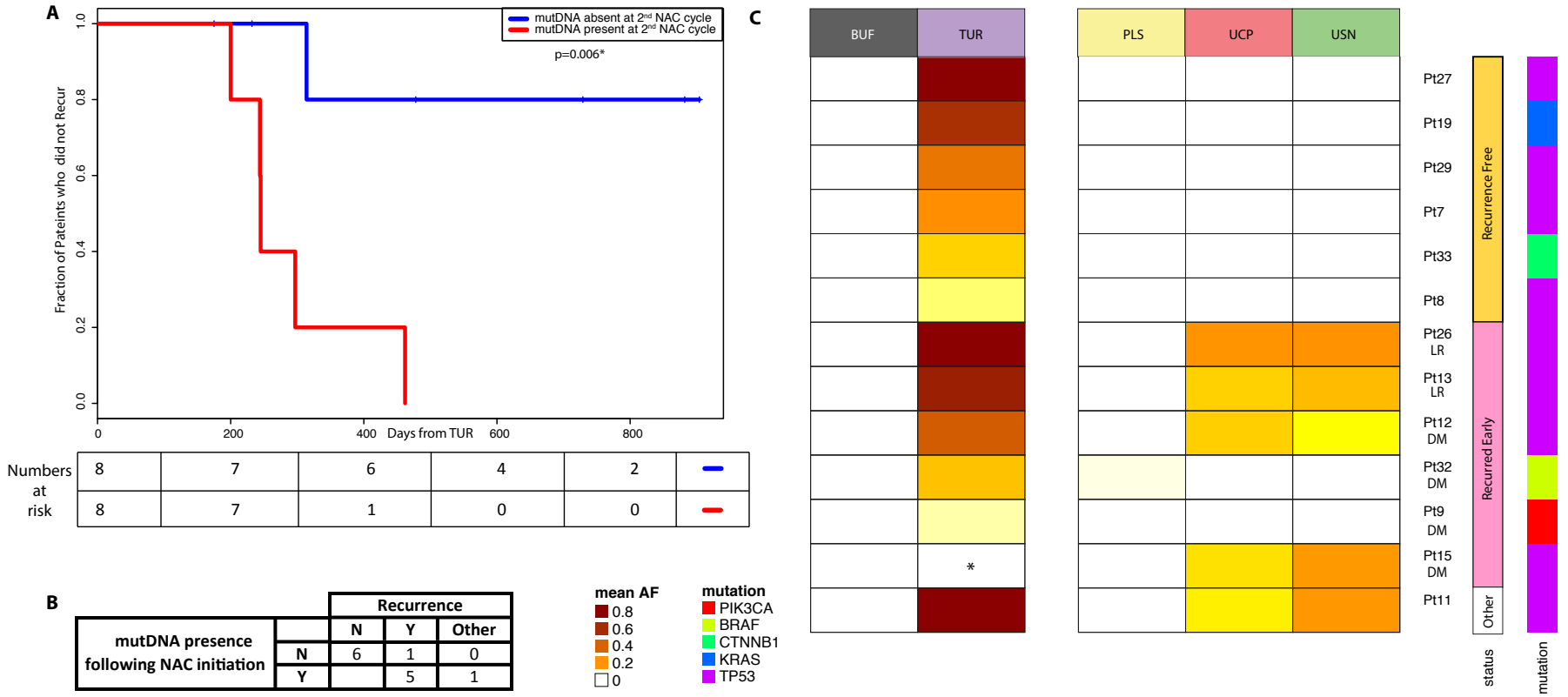



**Figure 3.10: Presence of mutDNA at the 2nd cycle of NAC predicts early recurrence in MIBC. A. Kaplan-Meier curve depicting time to recurrence from initial TUR.** We compared the rate of recurrence of patients with detectable mutDNA (red line) and undetectable mutDNA (blue line) in peripheral samples taken immediately prior to the 2nd cycle of NAC (i.e. 2-3 weeks after the initiation of NAC). MutDNA was detected in 5/6 patients who recurred and in 0/6 recurrence free patients. Median time to recurrence in patients with detected mutDNA was 293 days while in patients with undetected mutDNA the recurrence rate was low. Numbers at risk for the days from TUR are shown below the x axis. **B. Sensitivity and specificity for recurrence prediction.** Overall sensitivity and specificity were 83.3% and 100% with positive predict value and negative predictive values of 100% and 85.7% respectively. One "other" patient was excluded from recurrence analysis due to post-operative death. **C. Heatmap comparing SNV maximum AF across all sample types and recurrence states at this time-point.** Mutant allele fractions (mAFs) are represented by coloured cells ranging from white to scarlet as mAF increases. Patients are grouped by recurrence status, with recurrence type below, DM - distant metastasis, LR - Local recurrence. Generally, SNV mAFs are noticeably higher in USN and UCP as compared to PLS. There is a clear correlation between SNV mAF in peripheral samples and patient recurrence status. *TUR material was not available for patient 15.

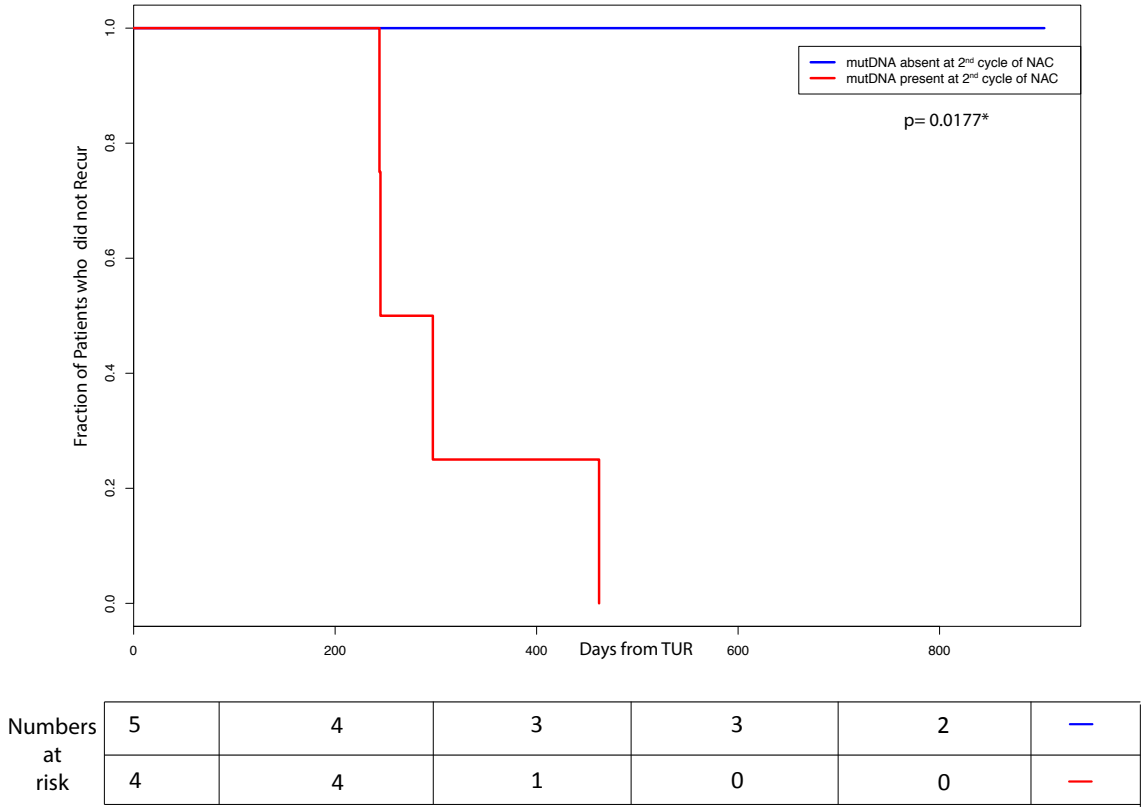



**Figure 3.11: Kaplan-Meier curve depicting time to recurrence from initial TUR in a subset of 9 patients who had *TP53* SNVs detected in their tumour sample.** All 4 patients who had *TP53* SNVs detected in their USN at the 2nd cycle of NAC (red line) recurred, with a median time of 271 days. The 5 patients who did not have a *TP53* SNV detected in their USN at the 2nd cycle of NAC (blue line) did not recur. When applied to MIBC patients with *TP53* SNVs detected in tumour specimens, the assay had a sensitivity and specificity of 100% (95% CI: 28-100% and 36-100%, respectively) for prediction of early recurrence. Numbers at risk shown below the x axis.

### 3.6.7 Comparison of peripheral sample types reveals that UCP and USN are enriched in mutDNA as compared to plasma

At 58 time-points all 3 sample-types were drawn simultaneously. Each of the 31 mutations was detected in at least one of the peripheral samples in this group. To identify the most informative peripheral sample type for mutDNA analysis in MIBC, each SNV detected at a single time-point was analysed as an independent variable. SNVs were detected most frequently in USN (34.5%, 49/142), UCP (27.5%, 39/142) and lastly plasma (9.9%, 14/142), Figure 3.12A. CNA analysis was conducted for the first and last time-points, for each patient where all three peripheral samples were analysed (35 time-points, patient 32 had only 1 time-point where all three sample types was drawn). When CNAs were detected, they were most frequently found in all three peripheral sample types. More CNAs were detected in USN and UCP than plasma but the small numbers precluded statistical interpretation Figure 3.12B.

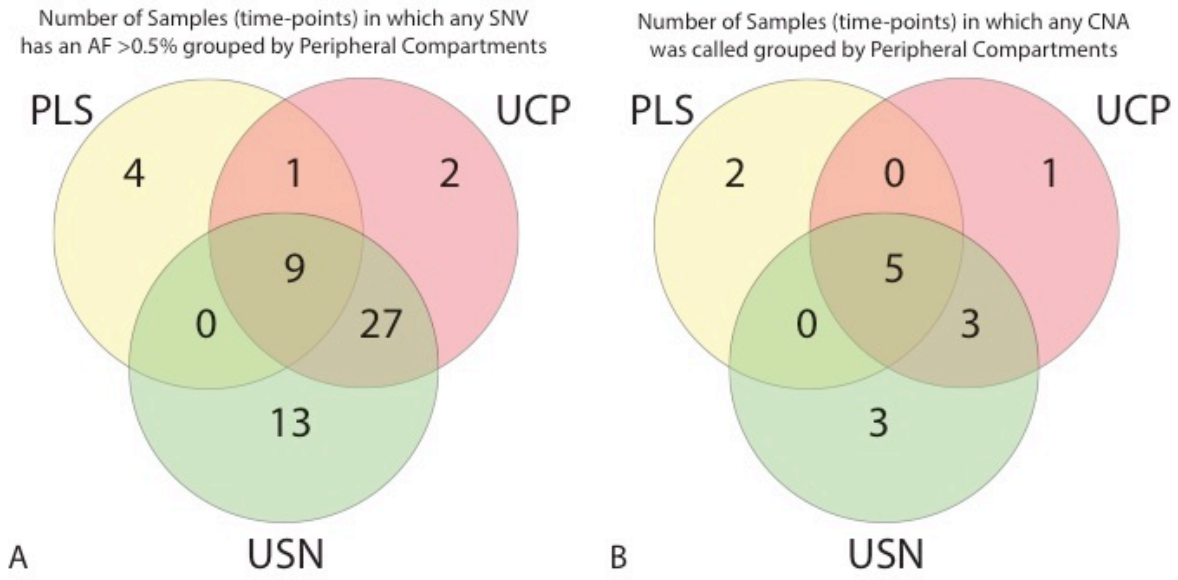



**Figure 3.12: Venn diagrams demonstrating that more SNVs (A) and CNVs (B) were detected in the urine, as compared to the plasma samples for time-points where all three sample types were collected.** The number of times SNVs or CNVs were detected in peripheral samples per time-point are depicted as Venn diagrams. CNAs were only assessed for the first and last time-points for each patient. 52 out of 56 SNVs and 12 out of 14 CNAs were detected in urinary samples. However for SNVs, 4 mutations/time-points were detected only in PLS, 2 in UCP and 13 in USN. For CNAs 2 mutations/time-points were detected only in PLS, 1 in UCP and 3 in USN, confirming that multiple sample analysis can improve mutDNA detection in MIBC.

Figure 3.13 is a box-plot of the raw SNV AF values, grouped by sample type. Mutant AFs were analysed for samples where SNVs were detected in one or more sample type(s) at that time-point without imposing detection criteria (Forshew *et al.*). Kruskal-Wallis testing demonstrated statistically significant differences in the AFs between sample types ($p= 4.527 \times e^{-12}$). Post-hoc Dunn-testing revealed significant differences between the AFs of UCP and plasma ($p=2.12 \times e^{-15}$) and USN with plasma ($p= 3.61 \times e^{-05}$). Mutant SNV AFs were not statistically different between USN and UCP (p=0.6975), Figure 3.13. However, no single peripheral sample type captured all of the SNVs that were detected across all the samples together, and for each sample type there were SNVs that were unique to it. Indeed, 4 of the events (individual SNVs detected in individual time-points) were detected only in plasma, 2 only in UCP, and 13 only in USN, Figure 3.12.

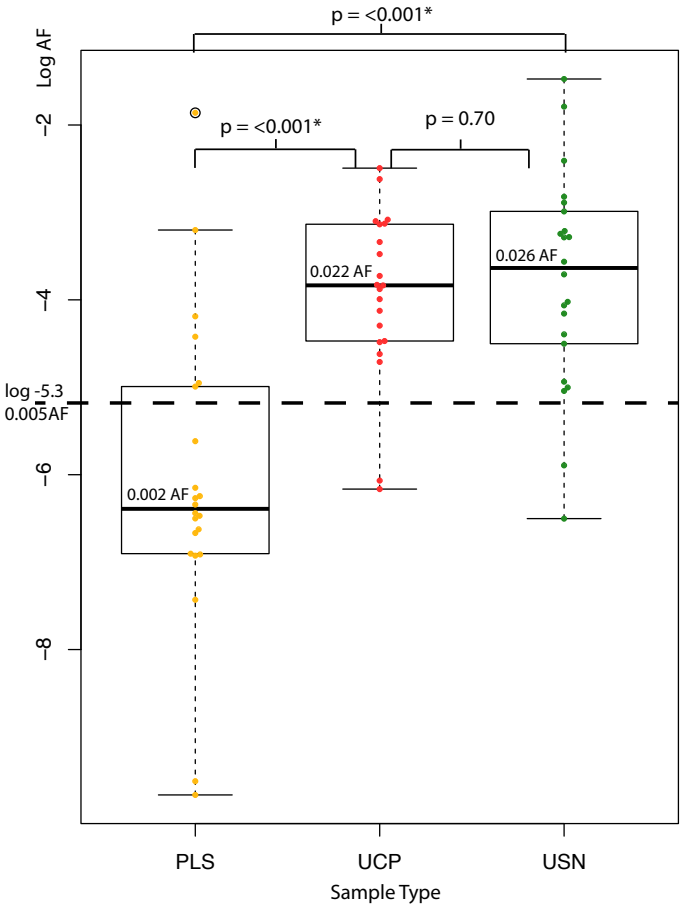



**Figure 3.13: Allele fractions of paired plasma and urinary samples taken from patients with MIBC undergoing NAC.** 86 time-points had paired sampling of all 3 peripheral sample types. In 22 cases a mutation was detected in one sample type, this and its paired sample AFs were plotted above with median AFs. The dashed line represents the technical threshold value of 0.5% AF. Kruskal Wallis testing revealed significant differences between the AF of urinary samples and the plasma but no difference between USN and UCP mAFs.

### 3.6.8 Comparison of CNA and SNV detection

In peripheral samples, though detection rates for SNVs and CNAs were similar, SNVs were not always detected in the same samples as CNAs. Therefore, SNVs and CNAs were compared for co-occurrence and whether there was a relationship between them. All peripheral samples taken from the first and last time-point that had one or more SNVs detected were included in the comparison (n=28). Figure 3.14 shows a bar chart of the samples ordered by the maximum AF (y-axis) and labelled by sample type (x-axis). Samples in which a CNA was detected are coloured red, whilst copy number neutral samples are coloured grey. The sensitivity of detection of CNAs using sWGS is considered to be 5-10% (Heitzer *et al.*)2011}, and it is noteworthy that above this threshold (as indicated by SNV AF) all samples had detectable CNAs. However, there were examples of CNA detection in samples where SNV AFs were low or were absent altogether (Figure 3.5 and Figure 3.14). This was apparent in all peripheral sample types.

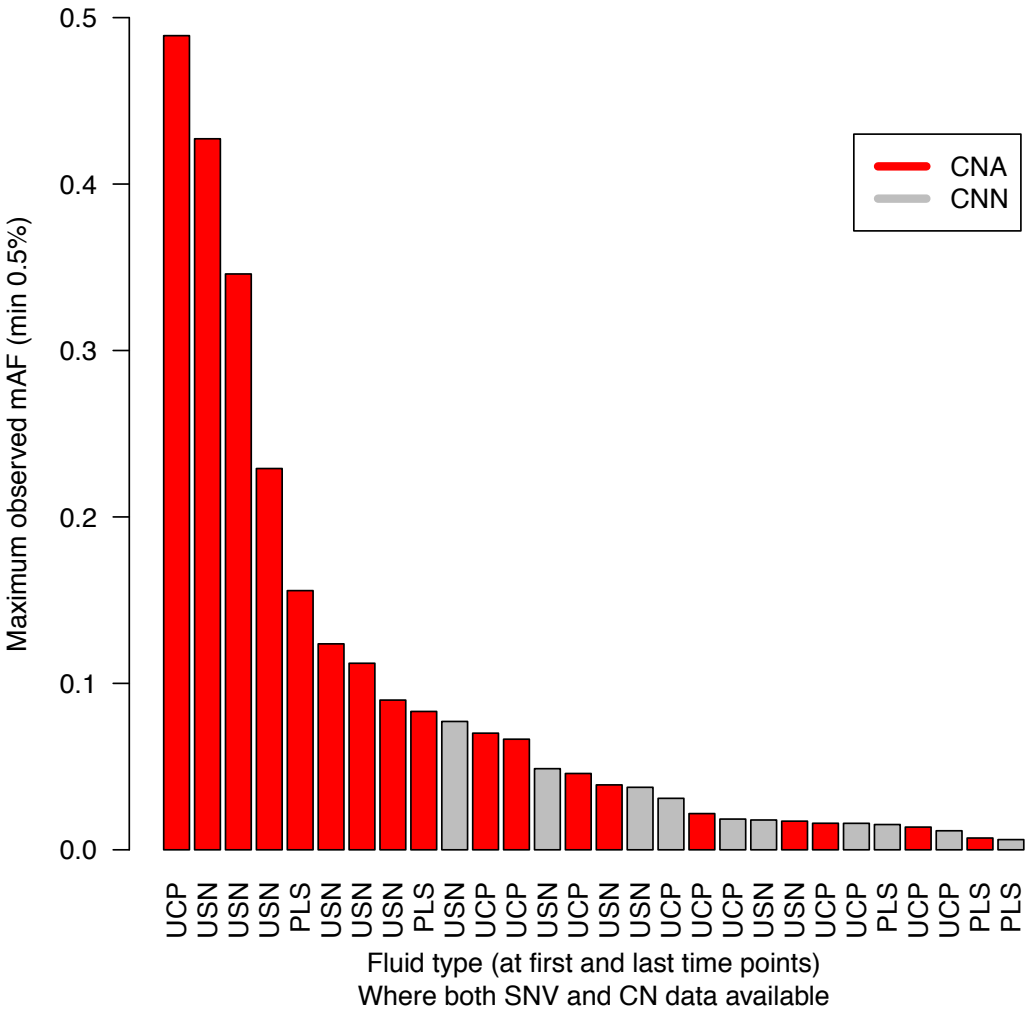



**Figure 3.14: Waterfall plot depicting the relationship between CNAs and SNV AF.**

This may be due to samples being primarily driven by CNAs and carrying only sub-clonal SNVs who's AF representation in plasma or urine are diminished as compared to tumours, or due to a clone having a particularly prominent and focal CNA that was therefore more easily detectable. Alternatively, this may be due to the ability of sWGS to interrogate a greater breadth of the genome than targeted re-sequencing approaches. This dataset demonstrates that sampling multiple body fluids using complementary techniques allows for more complete assessment of mutDNA.

### 3.6.9 Longitudinal analysis of mutDNA in peripheral samples of patients with MIBC

TAm-Seq was applied to analyse recurrent bladder cancer associated genetic events in serial USN, UCP and plasma peripheral samples (taken from NAC initiation through to its completion) to assess trends in mutDNA. Peripheral samples were collected longitudinally

from each patient over a median of 83 days (46-118 days, median of 15 samples/patient). CNA analysis was initially applied to the first and final time-points for all patients across all peripheral sample types.

Figures 3.15-7 show SNV AF kinetics for each patient grouped according to their recurrence status. Overall, the SNV AFs trend downwards and are undetectable after the initiation of NAC with the exception of the early recurrence group, where SNV AFs persist or even increase, most notably in the urinary compartments. Figure 3.18A shows a summary of longitudinal SNV AF kinetics for all patients distilled into 3 plots, one for each peripheral sample type. There are clear differences in the AF kinetics between peripheral sample types. For example, in patient 26, mutDNA levels at time-points 2 and 3 vary between UCP, where levels decrease, and USN, where levels increase. Furthermore, there are differences in the kinetics between patients. For example, for many patients, mutDNA levels fall during NAC with the exception of a few notable cases, e.g. patient 15 has increasing levels of mutDNA in USN samples.

Figure 3.18B similarly shows longitudinal mutDNA kinetics in 3 plots, one for each sample type. CNA analysis was performed routinely on the first and last time-point for each sample (x-axis). The y-axis demonstrates the Genome Wide Imbalance Score calculated by modelling the autosomal 1Mb bin read counts in samples against those in a control sample, and subsequent summation of the 5% most extreme residual values (i.e. greatest difference vs. copy number neutral), as described in the methods section (Section 3.4.4). Two predominant trends are observed; firstly, a series of patient samples that have high levels (>7) of CNA at their first time-point but much reduced levels at their final time-point. Of note, with the exception of patient 33, this series was enriched for patients that recurred early. The second trend observed involves ‘genome-wide imbalance scores’ that persist at low levels (<7) throughout NAC. Overall, the magnitude of CNA signals was higher in USN as compared to UCP, while levels were very low in PLS.

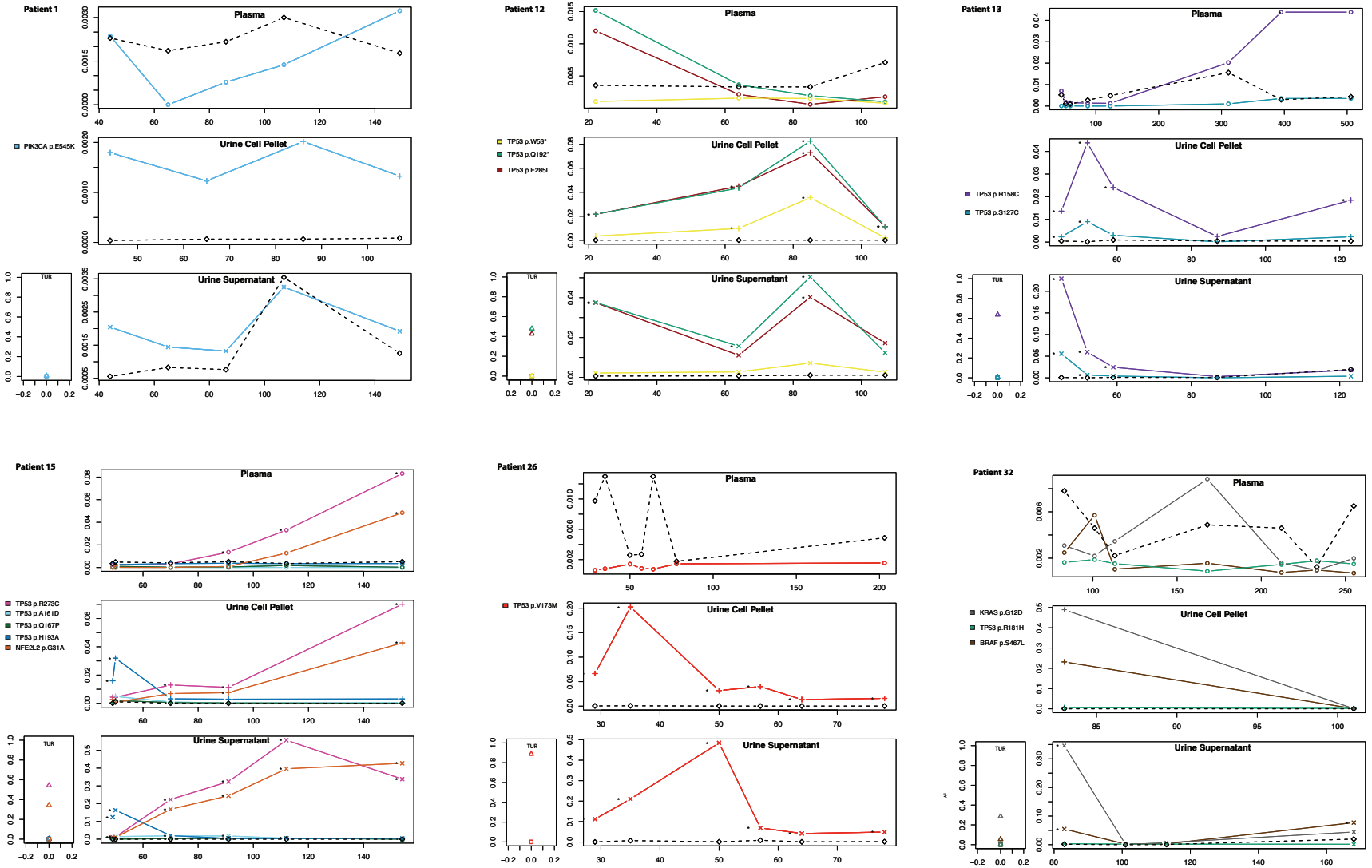



**Figure 3.15: SNV mutDNA kinetics for each peripheral sample type across the 6 patients who recurred early.** Generally all 6 patients had high levels of mutDNA (most often) in their urinary samples. Plots are grouped by patients into sets of four (with one plot for each peripheral fluid type and one for TUR and BUF samples). Each plot depicts days from TUR on the x-axis and mutDNA AF for SNVs at that time-point on the y axis. Data-points are colour coded according to the mutation ID, as per the figure legend. Notably, SNV kinetics of patient 9 were low in all sample types, including TUR indicating that the PIK3CA SNV that we track was not a component of the major clone.

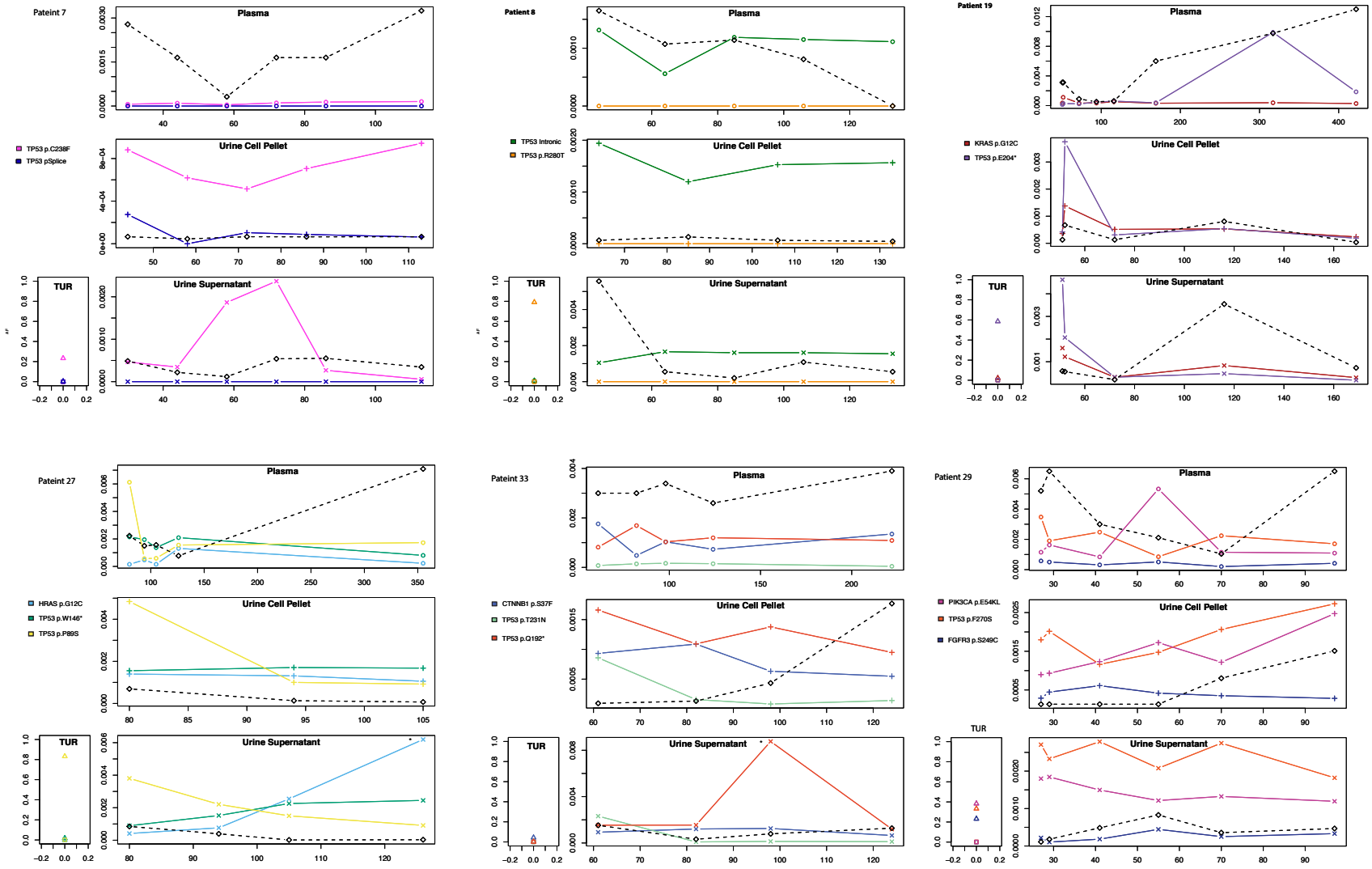



**Figure 3.16: SNV mutDNA kinetics for each peripheral sample type across the 6 patients who were free from early recurrence.** Kinetics demonstrate generally low levels of mutDNA in all patients regardless of sample type**.** The format of these figures match the format described in **Figure 3.15**.

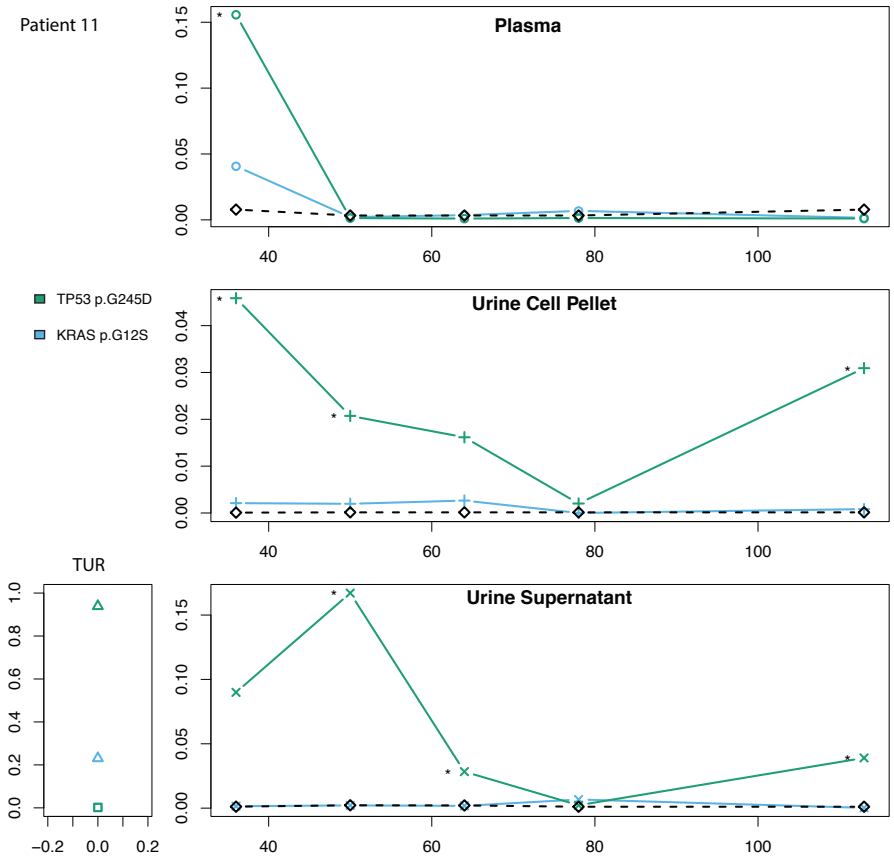



**Figure 3.17: SNV mutDNA kinetics for each peripheral sample type of patient 11, who died shortly after surgery from surgical complication.** The format of this figure matches the format described in **Figure 3.15**.

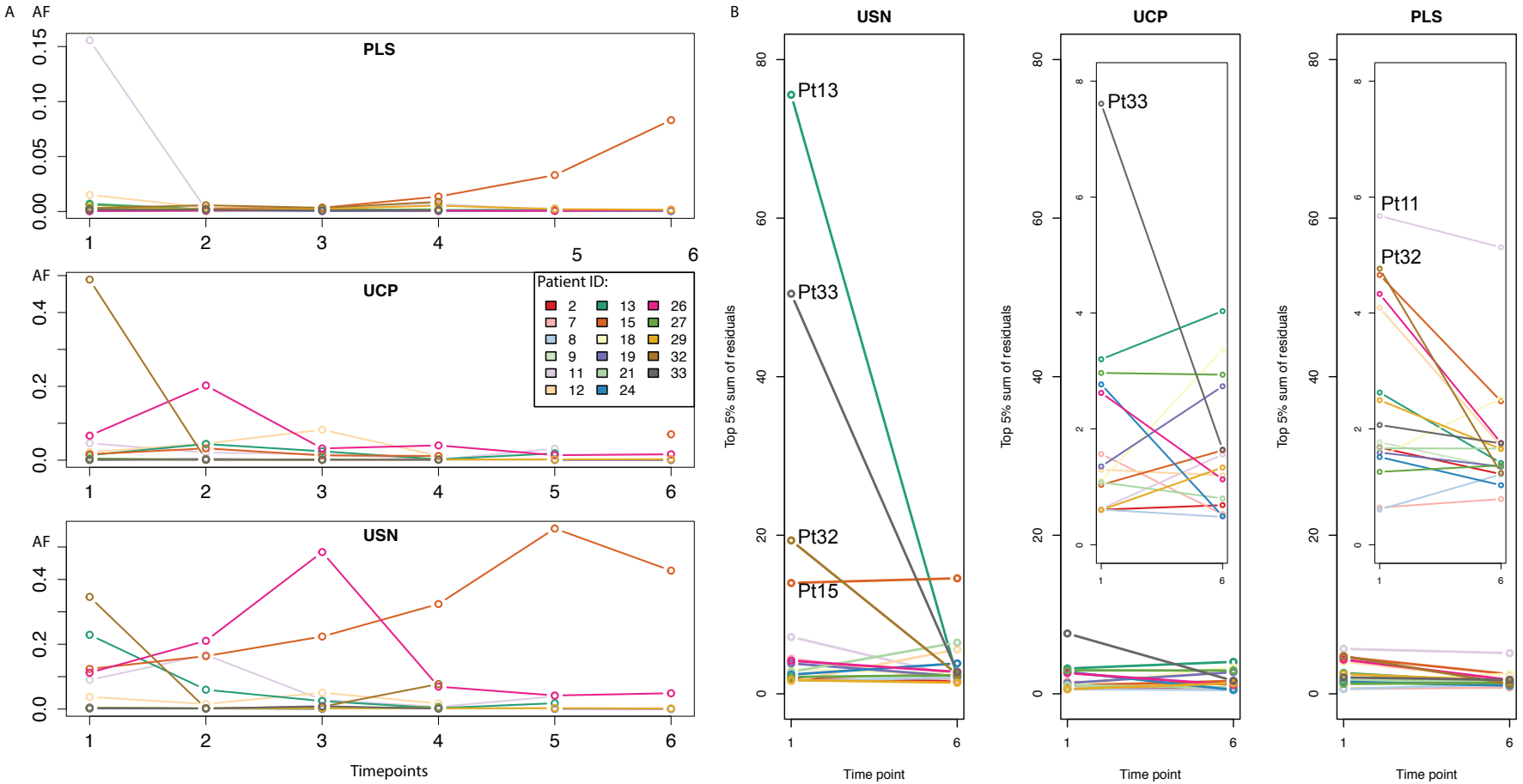



**Figure 3.18: Summary of longitudinal dynamics of patient specific SNVs and CNAs. A. Patient specific SNV kinetics across PLS, UCP and USN samples.** Longitudinal SNV AF kinetics are depicted in 3 plots, one for each peripheral sample type. Each plot depicts time-points on the x-axis. Only the SNV with the highest AF at a given time-point is plotted, with line colours corresponding to individual patients. **B. Changes in patient specific 'genome-wide imbalance scores' between first and last time-points, across peripheral sample types**. GWISs kinetics are depicted in 3 plots, one for each peripheral sample type. Each plot depicts time-points on the x-axis. GWISs were calculated as described in the methods section but, briefly, involved linear modelling of autosomal 1Mb bin read counts in samples against those in a control sample, and subsequent summation of the 5% most extreme residual values (i.e. greatest difference vs. copy number neutral). UCP and PLS GWIS plots are supplemented by an inset expanding the low imbalance score range.

For both SNV and CNA dynamics, the overall trend was a reduction in mutDNA over time during NAC (Figure 3.18). However, there were examples of persisting mutDNA, particularly in the urinary samples. Furthermore, there were patients who showed variation in mutDNA AFs over time. Figure 3.19 shows longitudinal SNV kinetics in a similar format to Figure 3.18A. There are clear differences in the AF kinetics between the peripheral sample types. Generally levels are low in PLS while AFs rise and fall dynamically in urinary specimens. For most patients, mutant allele fractions (mAFs) are low during NAC, however, mAFs that were considerably higher than the 0.005 AF detection threshold were more found in patients that recurred, see Figure 3.19. These general tends remained true when evaluating all SNVs detected in each patient.

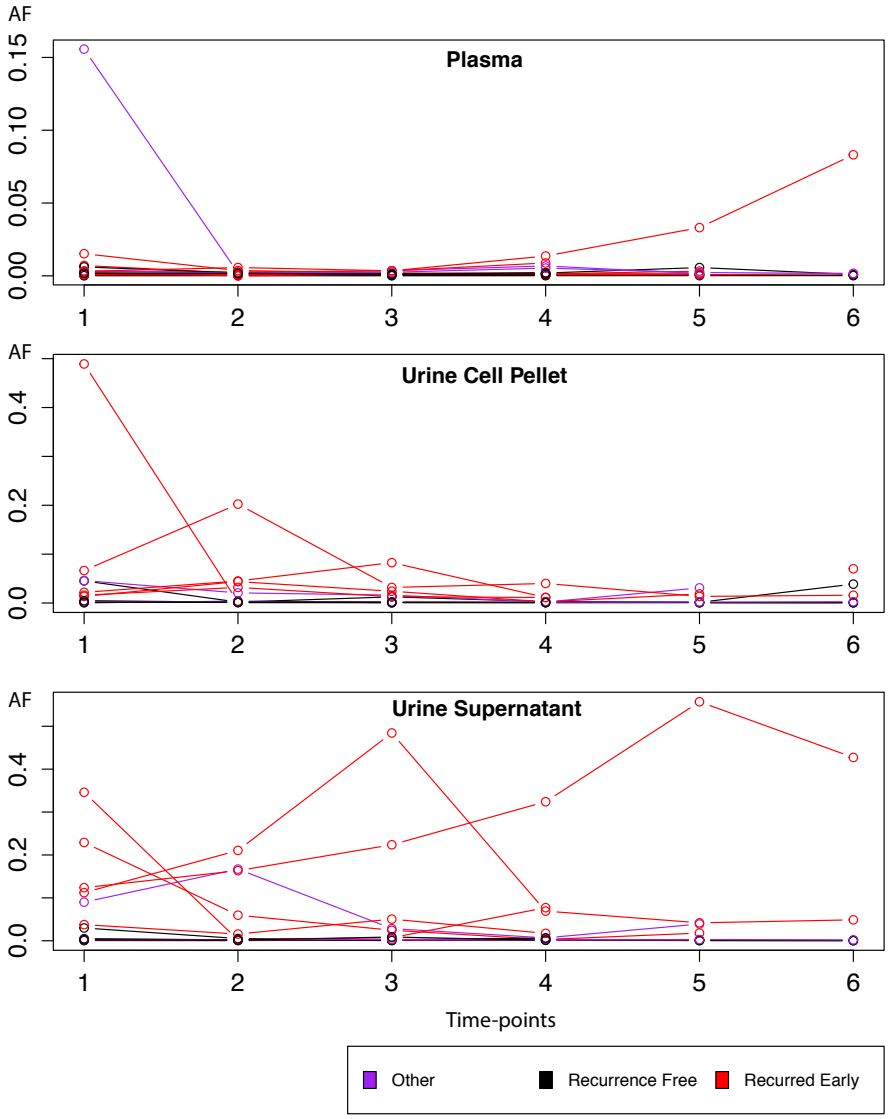



**Figure 3.19: Maximum mutDNA AF during NAC demonstrates differing kinetics in PLS, UCP and USN.** The format of these plots is similar to that of **Figure 3.18**. The 3 plots depict the maximum SNV AF at each time-point in PLS, UCP and USN samples for 13 patients with detected SNVs. Patients data points are colour coded by recurrence status.

### 3.6.10 Urinary mutDNA demonstrates tumour evolution during therapy

As opposed to assays targeting mutations detected in matched tumour samples, combined use of our disease specific assay and sWGS allowed the detection of *de novo* mutations. To determine additional examples of tumour evolution we performed sWGS analysis of samples taken between the first and last time-points in selected patients based on high SNV levels or interesting CNA profiles at the first time-point (Figure 3.5B). Indeed, in 5 patients there was evidence of dynamic tumour evolution during NAC, highlighting the strength of studying mutDNA in peripheral fluids as an alternative to traditional biopsy approaches (Murtaza *et al.;* Forshew *et al.*). All *de novo* mutations were detected in the urinary specimens and were not detected in the initial tumour specimen. There was no correlation between the detection of these private mutations and clinical outcome.

In patient 12, whilst mutDNA levels in plasma fell quickly following the initiation of NAC, urinary mutDNA levels remained high and reached a peak at 85 days after TUR. Additionally, a new nonsense mutation of *TP53* (W53*) was identified in urinary samples at this time-point (3.6% in UCP and 0.7% in USN), suggesting the emergence of a new clone (Figure 3.20A). Similarly overall CNA levels rose in parallel with SNV AFs and also demonstrated the emergence of a new CNA profile (including a novel focal amplification of *GRIN2A*. *GRIN2A* encodes an ionotropic glutamate receptor that regulates the influx and efflux of cations that are important in cellular migration and survival. Amplification of *GRIN2A* has previously been observed in MIBC and has been implicated as a cancer oncogene (Morrison *et al.;* Wei *et al.*, 2011). The short-term persistence of this clone was confirmed by sWGS analysis of additional urine samples from this patient (Figure 3.20B). Levels of all mutations ultimately fell, though some were detected at the last sample taken during NAC. Although a good response to NAC was initially reported for this patient, he developed brain metastases approximately 2 months following radical radiotherapy to the bladder.

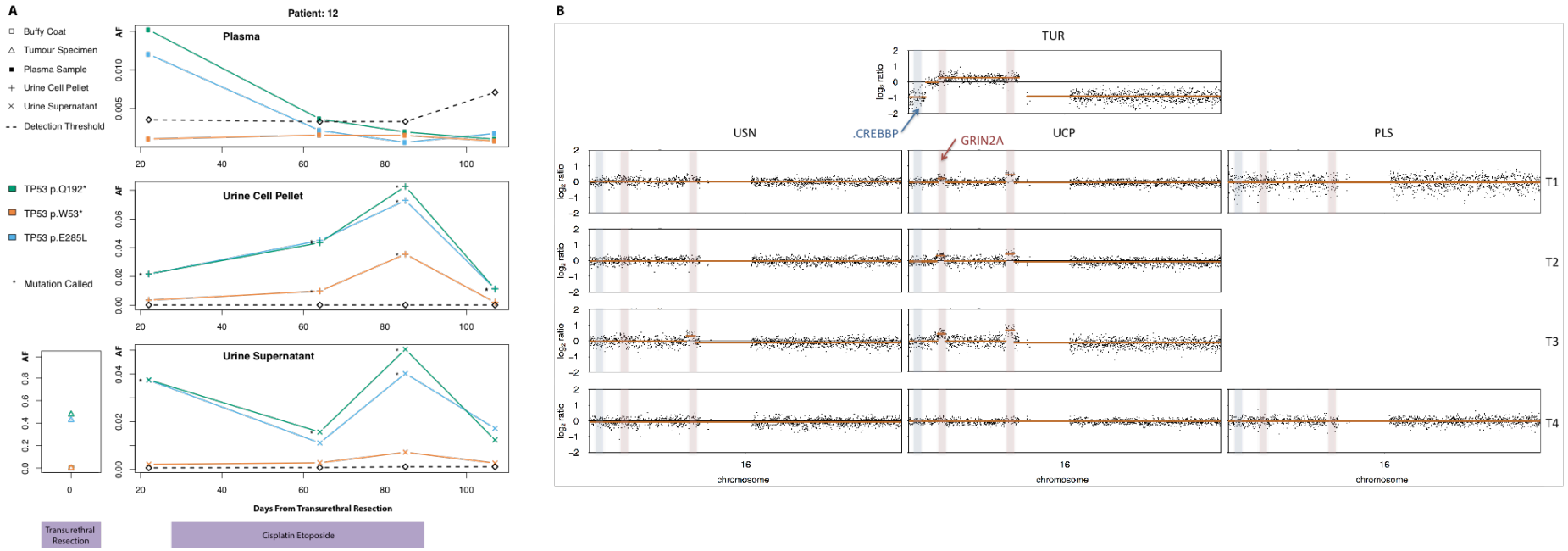



**Figure 3.20: MutDNA kinetics of patient 12 reveal tumour evolution on therapy. A. Kinetics of SNVs in longitudinally obtained peripheral samples from patient 12.** Changes in SNV AFs are plotted for PLS, UCP and USN from patient 12. SNVs in *TP53* (Q192* and E285L) are at their highest levels (as determined by AF) in PLS at time-point 1 but decrease below detection thresholds during NAC. Meanwhile, in UCP and USN, these mutations reach a peak AF at the third time-point (85 days after TUR). Coincident with this peak is the emergence of a third mutation of *TP53* (W53*). This mutation, which was not detected in the original TUR, reaches a modest AF of 3.6% in UCP and 0.7% in USN before receding at the final time-point (along with the other two *TP53* mutations). Despite its brief appearance, this third SNV indicates the emergence of a new clone under the selective pressure of NAC. **B. Evidence of changing clonal dynamics during NAC through CNA analysis.** Coincident with the emergence of *TP53* W53*, is the detection of a change in patient 12's CNA profile. At TUR, the genome-wide CNA profile consists of multiple amplifications and losses involving both large chromosomal regions (sometimes whole chromosomes) and focal areas. This is exemplified by chromosome 16 (plotted), where 16p contains neighbouring regions showing focal loss (region shown in blue - including the *CREBBP* gene (*20*)), and large-scale amplification respectively. In peripheral samples, the TUR CNA profile (chromosome 16 and genome-wide) is largely absent. However, focal amplifications of two regions on 16p are observed, one of which contains *GRIN2A* (regions shown in red). The amplitude of these CNAs reached their peak at time-point 3 (in UCP) before receding at time-point 4. It is unclear if the focal amplification of GRIN2A is present in USN (low level CNAs) and PLS (time-points missing). Combined, the SNV and CNA data point to the emergence of a clone containing a *TP53* W53* SNV and a focal amplification of *GRIN2A*. It is unclear whether this clone has any influence on ultimate disease course in this patient who ultimately developed brain metastases after radical radiotherapy. Nonetheless, our data demonstrate the potential of mutDNA (in this patient, primarily derived from UCP) in tracking tumour evolution during therapy.

In another example, patient 15 had T3N1 disease that recurred 138 days after cystectomy. The SNV and CNA profiles observed in her urine specimens suggested tumour evolution in response to surgery and/or NAC. Specifically, a clone characterised by a *TP53* H193A SNV and a focal amplification of *YAP1* was dominant at the pre-NAC time point. This sample also carried *TP53* R273C and *NFE2L2* G31A SNVs but they were detectable at low AFs (Figure 3.21A and B) and likely represent a minor clone. During NAC, the initially dominant (presumably NAC sensitive) *TP53* H193A, containing clone receded, whilst the minor clone containing *TP53* R273C and *NFE2L2* G31A SNVs and a focal loss of *CDKN2A* became dominant in all peripheral fluid samples. Radical cystectomy was carried out 155 days after TUR and 106 days after the initiation of NAC. We obtained DNA from the cystectomy sample and carried out TAm-Seq and sWGS on it. Of note, we found that the clone containing *CDKN2A* loss and *TP53* p.R273C and *NFE2L2* p.G31A SNVs was present at high levels, whilst there was no evidence of the clone containing *YAP1* gain and the *TP53* p.H193A SNV (Figure 3.21A and B). The similarity between peripheral samples obtained during NAC and the subsequent cystectomy samples was confirmed by modelling the linear relationship between CNA profiles of these samples (Figure 3.21C). We also carried out additional sWGS of intermediate time-points for this patient and confirmed gradual loss of the pre-NAC CNA profile below our detection threshold. Together this data points to ‘on-therapy’ tumour evolution under the selective pressures of NAC, as indicated by apparent changes in the dominant clone and its respective AF (Figure 3.21D). Whilst most noticeable in USN, this evolution was also evident in matched UCP samples, albeit at lower levels. Plasma mutDNA analysis alone would not have demonstrated evolution in this case, although it did reveal the emergence of the later clone containing *CDKN2A* loss and *TP53* p.R273C and *NFE2L2* p.G31A SNVs, Figure 3.22. This is possibly due to low mutDNA levels at the early time-points, which may in turn be due to spatial differences in the clones resulting in different representation of shed DNA in the plasma and urine.

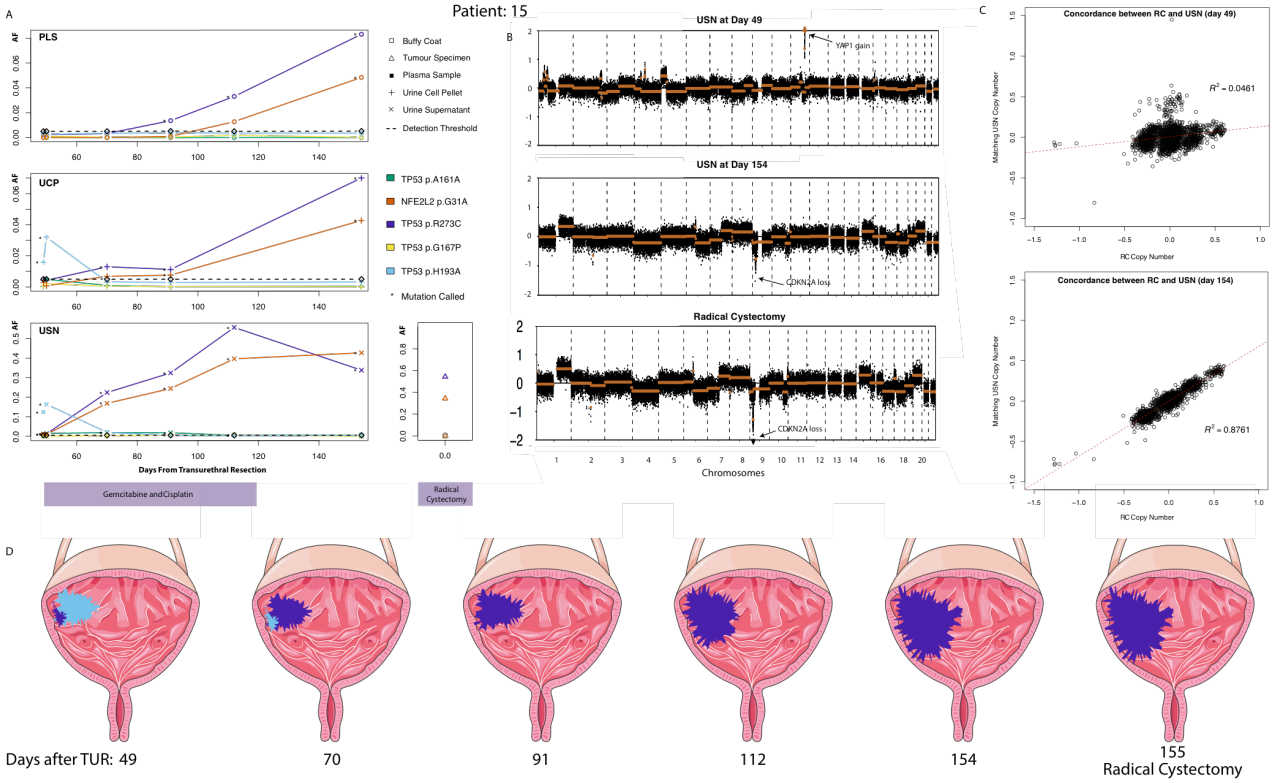



**Figure 3.21: mutDNA demonstrates on therapy tumour evolution. A. SNV plots for patient 15 demonstrate tumour evolution.** SNV analysis of pre-NAC urinary samples revealed a *de novo TP53* H193A mutation (light blue), whilst *TP53* R273C (purple) and *NFE2L2* G31A (orange) SNVs were only observed at low AFs in USN samples (1% and 1.1% respectively). During NAC, the clone containing *TP53* R273C and *NFE2L2* G31A SNVs appears to grow considerably, whilst the *TP53* H193A SNV containing clone recedes to become undetectable at later time-points. This profile was mirrored by the cystectomy sample. **B. CNA profiles demonstrate tumour evolution.** Marked CNA changes were observed in urine sample at pre-NAC (including *YAP1* focal amplification). This CNA profile differed from that seen in later time-points. At time-point 6, the CNA profile resembles one seen at cystectomy. **C. Concordance of cystectomy and USN CNA profiles during NAC.** We generated a linear model by fitting autosomal 1Mb bin read-counts in the cystectomy sample against those in peripheral samples. Initial USN CNA profiles are discordant with the cystectomy sample ($R^2$=0.0461). Subsequently a concordant CNA profile emerges ($R^2$=0.8760), mirroring the SNV results. **D. Longitudinal mutDNA analysis suggests on-therapy tumour evolution.** We used the changing SNV and CNA profiles to suggest a clonal evolution paradigm in patient 15. (Images adapted from Servier Medical Art).

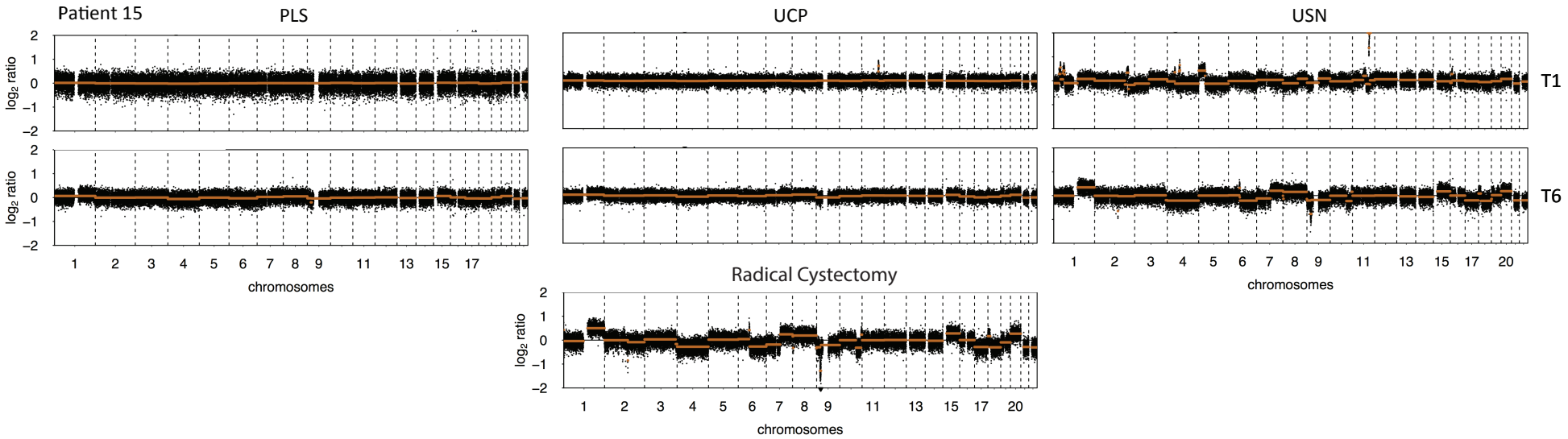



**Figure 3.22: Patient 15 CNA profiles of all three peripheral samples taken pre-NAC (T1), prior to Radical Cystectomy and lymph node dissection (T6) and from the radical cystectomy specimen.** At T1 a CNA profile consisting of multiple amplifications and losses is obvious in the USN sample. Amongst these is a focal gain of *YAP1* on chromosome 11 which is also called in the matching UCP sample. At time-point 6, a completely different CNA profile is observed in USN (including focal loss of *CDKN2A* on chromosome 9). The UCP and PLS samples at this point are largely copy number neutral (though may show evidence of low-level *CDKN2A* loss). The cystectomy sample contains a similar CNA profile to that of USN T6, with focal *CDKN2A* loss and amplification of chromosome 1q. The dynamics suggested by these profiles match those observed through SNV analysis at the same time-points. Combined, this data points to changing clonal dynamics, with the emergence of a new clone following initiation of NAC. Simultaneously, the clone that was dominant at T1 recedes, suggesting that it may be sensitive to NAC. Indeed, of all patients considered in this study, patient 15 has the highest SNV AF at the final time-point (**Figure 3.18**) and examination of their clinical notes confirmed that they did not respond to therapy and later recurred. Therefore, mutDNA analysis of peripheral samples (here, in particular USN) can guide tracking of tumour evolution and disease course.

### 3.6.11 SNP analysis suggests sample crossover during the experiment is unlikely

Amongst the ~5Kbps interrogated by the bladder specific panel, there were 90 known SNP positions. Following TAM-Seq of the samples, AF's for each were tabulated with SAMtools (Li *et al.*, 2009) and SNP positions were removed if that position was found to be uninformative, i.e. if the AF was consistent throughout all of the samples as this SNP position in our cohort. Seven SNPs were found to be informative for comparison. Further filtering was performed based on the median sequencing depth at all positions. Sample coverage was plotted for each sample (see Figure 3.23) and 6 samples with a median coverage of <500x were removed from further analysis. The resulting AF data were plotted together with median coverage using 'Heatmap' function in R in Figure 3.24.

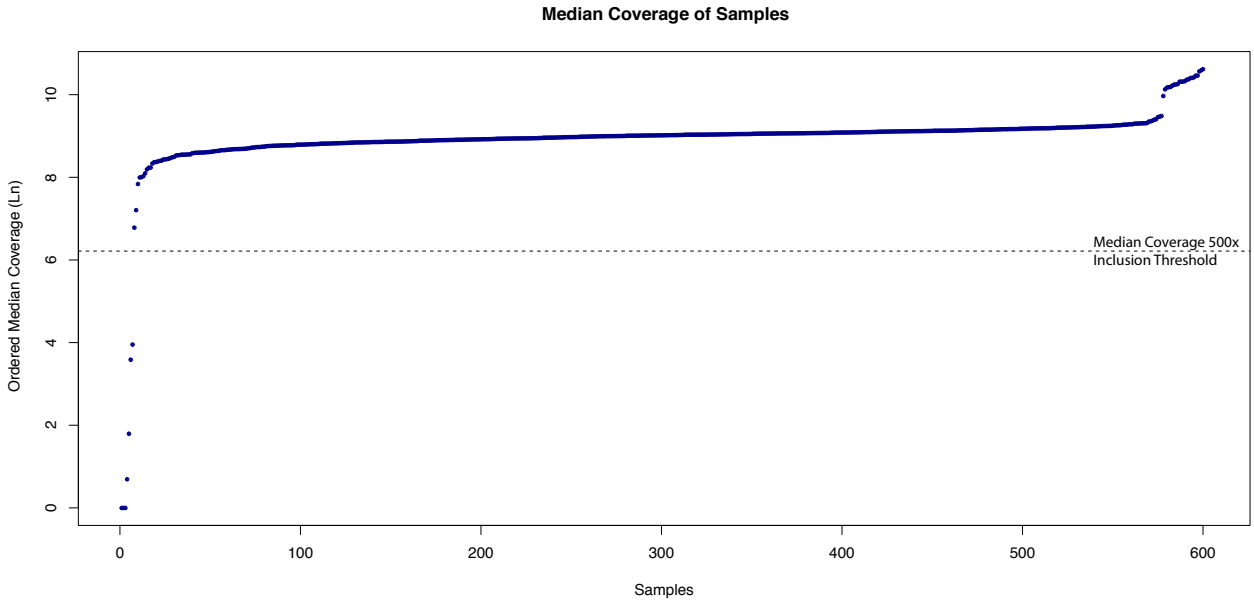



**Figure 3.23: Median coverage for all samples included in SNP analysis.** The y axis is scaled with natural log and the dotted line represents the coverage the threshold of 500x which was used to remove poor quality data.

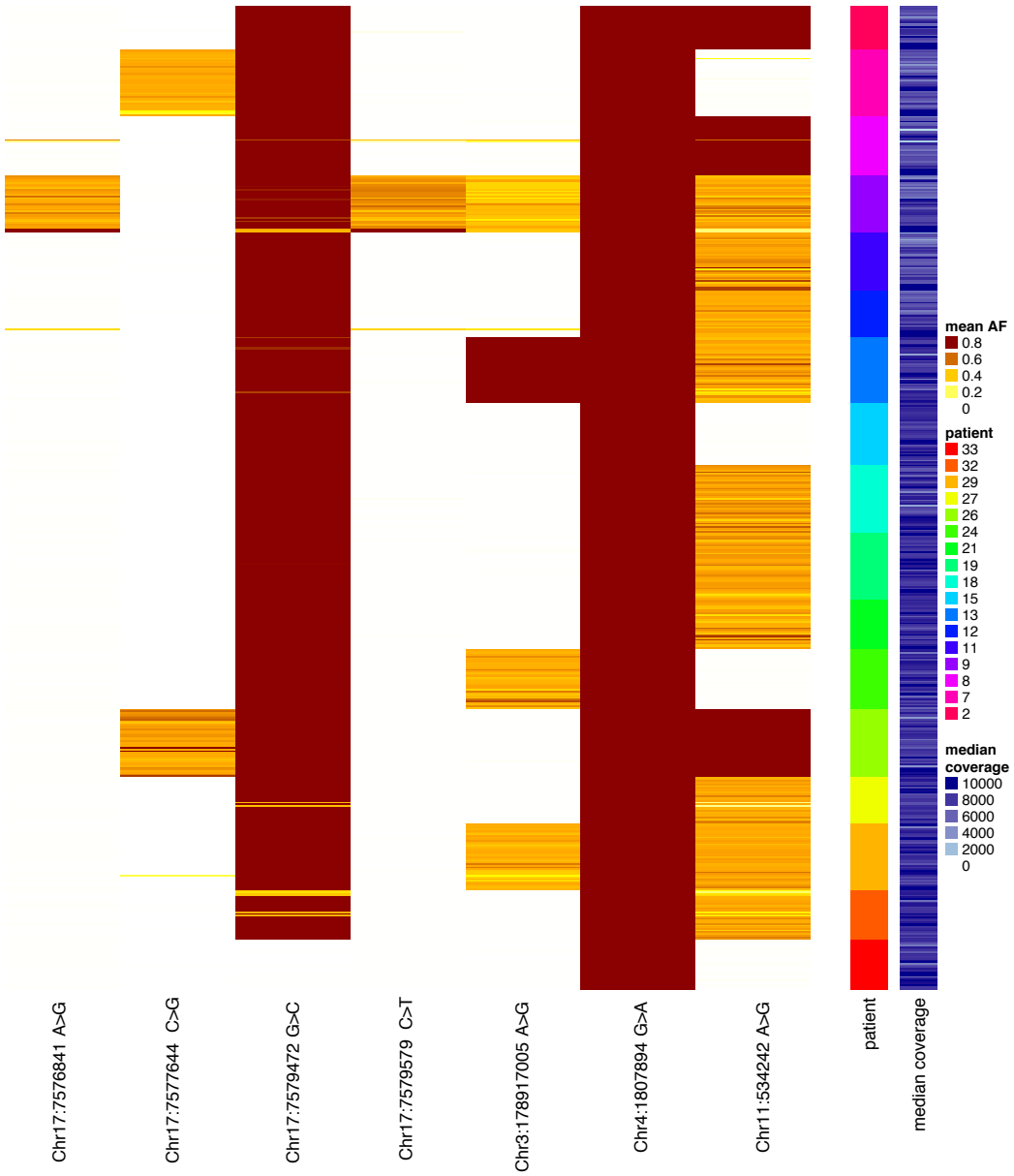



**Figure 3.24: SNP analysis demonstrating concordance of SNP genotypes for samples taken from the same patient.** The analysis of 7 SNPs (that were included in our TAm-Seq panel) revealed that most patients have unique SNP profiles. Manual clustering of SNP profiles according to patient ID showed that SNP profiles from samples from an individual patient match other samples taken from the same patient and differ from those taken from other patients.

Informative SNP analysis revealed unique profiles for the following patient IDs: 2, 7, 8, 9, 13, 15, 24, 26, 29 and 33. Patients 11 and 12 had an indistinguishable SNP profile from each other, as did patients 18, 19, 21, 27 and 32. However, the SNP profile of the later group of patients was distinguishable from that of former group of patients. Figure 3.24 demonstrates that samples collected from the same patient had identical SNP profiles to each other, indicating that sample mix up was unlikely. It must be noted that sample mix up between patients with an identical SNP profile or between time-points of the same patient cannot be ruled out. The likelihood of a sample being mistaken for another patient's

is however unlikely due to the low probability of it being mixed up in exactly the same position as another sample with the same SNP profile.

### 3.6.12 MutDNA analysis can be valuable in multiple disease settings of urothelial cancer

To explore the applicability of mutDNA detection in other urothelial cancer settings we used TAm-Seq with the bladder primer panel and sWGS to analyse mutDNA levels in 3 additional patients: Patient 1 had Urothelial Cell Carcinoma (UCC) of the distal ureter, patient 10 had metastatic UCC of the bladder and patient 30 had UCC of the renal pelvis. For patient 1 tumour DNA was extracted from biopsy tissue, for patient 10 DNA was extracted from TUR material prior to the diagnosis of metastatic disease, and for patient 30 DNA extraction was not possible due to the small amount of material obtained from the renal pelvis biopsy. All patients were treated with MVAC (methotrexate, vinblastine, doxorubicin and cisplatin). In addition, patient 1 had radical nephrouretectomy with lymph node dissection (NU + LND), see Table 3.7.

| Pt IDs | Age | Sex | Cancer type | No. of Samples Collected | | | | | Grade | Stage | Treatement |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Tumour DNA | BUF | PLS | UCP | USN | | | |
| 1 | 35 | M | ureteric UCC | 1 | 1 | 6 | 6 | 6 | 3 | T3N0 | NAC, NU + LND |
| 10 | 35 | M | metastatic UCC | 1 | 1 | 7 | 7 | 7 | 3 | T3N2M1 | Chemotherapy |
| 30 | 67 | M | renal pelvis UCC | 0 | 1 | 3 | 2 | 2 | 3 | T2N1 | Chemotherapy |

**Table 3.7: Patient demographics for additional patients are shown in a similar format to Table 3.2.**

For patient 1, SNVs in *TP53* and *KRAS* that were identified in tumour DNA, were detected at low levels (AFs <5%) in different peripheral samples at different time-points during treatment (Figure 3.25A). For patient 10, the CNA profile observed in the tumour (including focal loss of *CDKN2A* and *CREBBP*) was observed in all peripheral sample types taken prior to starting chemotherapy, albeit at lower levels in plasma. MutDNA levels, as inferred by the CNA profile, gradually decreased to below the detection threshold of sWGS, first in plasma, and then in UCP and USN during treatment (Figure 3.25C). For patient 30 the tumour biopsy did not yield sufficient material for DNA extraction. We detected SNVs in *TP53* and *PIK3CA*, as well as CNAs (including focal amplifications of *MYCL1, E2F3/SOX4* and *PPARG*, and focal loss of *CDKN2A*) in all peripheral samples. SNVs were detected at consistently high AFs in the urinary specimens but at moderate levels in the plasma. Matching CNA profiles were seen only in the urinary samples, where they persisted between the two time-points analysed (Figure 3.25B). These preliminary findings suggest

that analysis of mutDNA in body fluids may be useful for disease monitoring in multiple urothelial cancer settings.

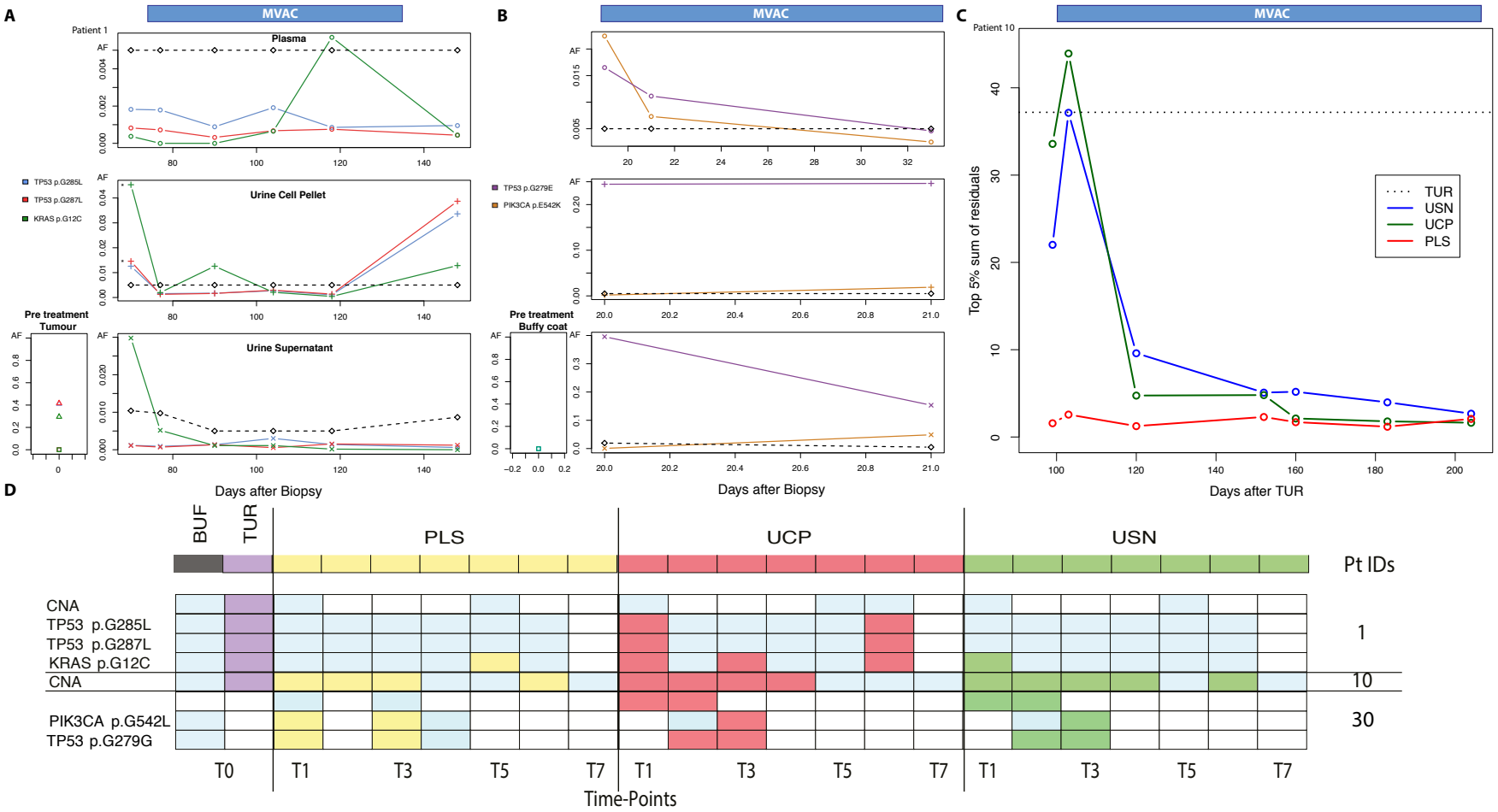



**Figure 3.25: mutDNA kinetics for patients with a broader spectrum of urothelial cancers.** The format of these figures match the format described in **Figure 3.15, Figure 3.18 and Figure 3.5B. A.** For patient 1 (ureteric UCC), SNV kinetics differ between the urinary component and the plasma, with initially moderate levels in the former that decrease during NAC before re-appearing in the later time-points. PLS samples have generally low mutDNA kinetics throughout with low level mutDNA detected at 118 days following tumour biopsy **B.** For patient 30 (renal pelvis UCC), clonal dynamics differed across the 3 sample types, in samples collected over the period of 1 week. TP53 p.G279E mutDNA levels in PLS and USN levels declined whilst remaining stable in UCP samples. However, the clone containing the PIK3CA p.E542K mutation increased in the urinary components whilst decreasing in the PLS samples. **C.** mutDNA kinetics based on a genome wide imbalance score (GWIS) in patient 10 (metastatic UBC) demonstrate concordant kinetics in the urinary components, with an initial spike in levels soon after the initiation of NAC and a subsequent decrease during NAC. **D. Grid depicting mutDNA detection across the 3 additional patients and time-points**

## 3.7 Original Contributions to Knowledge

1. DNA obtained from urine (UCP and USN) had a higher number of SNVs detected than plasma and a statistically significantly higher AF.
2. mutDNA analysis at early time-point during NAC could not predict pathological response to NAC but importantly, is associated with recurrence.
3. Tumour evolution can be demonstrated from urinary mutDNA and we describe the first example of monitoring tumour evolution through urinary mutDNA

This data suggests that mutDNA has promise as a biomarker to predict recurrence in MIBC and may be useful in monitoring tumour evolution.

# CHAPTER 4: ctDNA DETECTION IN LOCALISED PROSTATE CANCER

## 4.1 Synopsis

Prostate cancer is the most common cancer in UK men and its incidence is rising. Currently, men diagnosed with prostate cancer have repeated Prostate Specific Antigen (PSA) measurements to monitor their disease activity. Unfortunately, PSA is non-specific and cannot distinguish between indolent and aggressive disease, therefore repeated biopsies are often required for active surveillance of prostate cancer, conferring increased morbidity. Men with localised prostate cancer often have indolent disease that will not affect their lifespan and are over-treated and over-investigated. There is therefore a need for a biomarker that can accurately distinguish between indolent and aggressive cancer. ctDNA analysis may be a good candidate as samples can be peripherally acquired allowing for multiple sampling and may be able to highlight aggressive cancers through high ctDNA levels or through changes in mutational signatures. Therefore ctDNA analysis could be a useful tool to monitor men with localised prostate cancer.

I designed and validated a custom prostate cancer gene panel, assessing whole exons or hotspots of 15 genes commonly mutated in prostate cancer. From prostatectomy FFPE tissue of 19 men with localised prostate cancers the number of SNVs identified by this panel was low. However, in a separate group of 19 men who had metastatic prostate cancer, 10/19 had *TP53* SNVs indicating that certain mutations may predispose to aggression. Analysis of *TP53* in tissue samples taken at the time of prostatectomy, demonstrated the presence of *TP53* SNVs identical to those found later in metastatic clones in 6/6 men. Furthermore the same *TP53* SNVs, were detectable in pre-operative plasma samples from 2/3 men. To investigate whether mutations from more than one clone were represented in plasma, I developed MRD-Seq. When applied to 2 men with multifocal localised prostate cancer, mutations from multiple clones were detectable but clonal representation in the plasma was variable and overall levels of ctDNA was low.

This proof of principle study suggests that aggressive mutations are present and detectable in the plasma at the localised stage. Furthermore, that ctDNA analysis may allow the non-invasive monitoring of individual clones in multifocal prostate cancer. These results warrant further investigation. If corroborated, in the future ctDNA analysis could help monitor men on active surveillance and identify aggressive changes at an early stage.

## 4.2 Publications arising from this work

*Work presented in this chapter was described in my first year report and parts were published in Nature Communications. Further works described here are being prepared for submission to Clinical Biochemistry. The text is therefore excerpted from my first year report and related publications except to extend explanations, display updated data and provide more in-depth references.*

M.K.H. Hong, G.Macintyre, D.C. Wedge, P. Van Loo, **K. Patel**, S. Lunke, L. B. Alexandrov, C. Sloggett, M. Cmero, F. Marass, D. Tsui, S. Mangiola, A. Lonie, H. Naeem, N. Sapre, P. M. Phal, N. Kurganovs, X. Chin, M. Kerger, A. Y. Warren, D. Neal, V. Gnanapragasam, N. Rosenfeld, J. S. Pedersen, A. Ryan, I. Haviv, A. J. Costello, N. M. Corcoran & C. M. Hovens. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. Nature Communications. 6605:6. http://dx.doi.org/10.1038/ncomms7605

## 4.3 Aims

My primary objective was **to explore whether mutant cfDNA can be an informative marker for prostate cancers**. I collected samples from patients with localised prostate cancer who attended prostate clinics, using previously granted ethical approval (MREC:01/4/061, LREC:11/H0311/2 and LREC:03/018) and through international collaborations to investigate:

1. ***TP53* SNV analysis as a biomarker for aggression:** I investigated the ability of *TP53* detection during localised disease to predict future malignant potential using cohorts of men with well documented clinical outcomes and 3-5 year follow up.
2. **Whether multi-focal prostate cancer clones can be tracked at the localised stage:** Through international collaboration, I accessed multi-regional sequencing data of men with localised prostate cancer and designed assays to investigate whether these mutations can be detected in pre-operative blood samples.

## 4.4 Methods

### 4.4.1 Proof of Principle Analysis

Pilot analysis was retrospectively performed on 12 patient samples collected under the Prostate Cancer: Mechanisms of Progression and Treatment study (PrompT), MREC/01/4/061. Twelve cases were chosen as the initial panel was intended to detect SNVs in ~15% of cases, section 4.4.1.1. All men had previously undergone robotic assisted laparoscopic prostatectomy for prostate cancer and were selected due to large tumour size. Prostate tumour specimens from FFPE blocks and cores were taken to determine tumour grade and stage by Dr Anne Warren (Consultant Histopathologist), as previously described (Warren *et al.*).

FFPE slides were cut (6µm) and macro-dissection was performed, using 6-8 sections for tumours with an area of >1cm$^2$, or 10-12 sections for tumours with a surface area of <1cm$^2$. DNA was extracted from FFPE samples using the QIAamp DNA FFPE Tissue Kit (Qiagen 56404) using the manufacturers recommended xylene-extraction protocol, with the following alterations from the Rosenfeld group Standard Operating Procedure (SOP): 1) following addition of buffer 180µl of ATL, the sample was incubated for 15 minutes at 95°C; 2) the sample was allowed to return to room temperature before the addition of Proteinase K and continuation of the manufacturers protocol; 3) the elution step was adjusted to use only 50µl of buffer AVE and the elution step was repeated with final eluate. FFPE extracted DNA was quantified using a florescence-based DNA quantification method (Qubit, ThermoFisher).

Plasma cfDNA was extracted from 1ml of blood plasma using the QIAamp Circulating Nucleic Acid kit (55114). Manufacturers' instructions were followed with the following additional steps: 1) a known quantity of 'spike-in' *Xenopus Tropicalis* DNA was added to each sample and to a no extraction control sample to estimate yield; 2) following elution of DNA, the eluate was re-applied to the QIAamp mini column membrane and incubated for a further 3 minutes before final elution, as per the Rosenfeld SOP.

TAm-Seq was used to interrogate tumour samples for Single Nucleotide Variants (SNVs) as described by Forshew *et al.* (Forshew *et al.*) and in Section 2.4.5. In brief, up to 50ng of DNA was used as input for the TAm-Seq protocol. Each sample was 'pre-amplified' using a limited number of PCR cycles in a multiplexed reaction (with a pool of primers). 'Pre-amplified' samples were then aliquoted into multiple singleplex PCR reactions using an

Access Array (Fluidigm) microfluidic platform, before PCR-based barcoding, pooling and Illumina Sequencing (on MiSeq or HiSeq platforms).

#### 4.4.1.1 Initial Primer Panel

For initial proof-of-principle studies in localised prostate cancer I used a previously published primer panel that targeted genes commonly mutated in several cancer types (Forshew *et al.*). This panel included hotspot mutations in *PIK3CA*, *EGFR*, *BRAF* and *KRAS* and all exons of *PTEN* and *TP53*, Appendix A-5. TCGA data suggested that this panel could detect SNVs in ~15% of localised prostate cancer cases (Cerami *et al.*, 2012; Gao *et al.*, 2013; Cancer Genome Atlas Research, 2015), see Table 4.1.

| GENE | TARGETING | MUTATED CASES (%) | PROPORTION OF PROSTATE CANCER WITH ALTERATIONS |
|---|---|---|---|
| PIK3CA | p.E545, p.G545 & p.HC1047 | 1.8 | |
| EGFR | 10/28 exons | 1.2 | |
| BRAF | p.V600 | 2.4 | |
| KRAS | p.G12, p.G13 & p.Q61 | 0 | |
| PTEN | all exons | 2.7 | |
| TP53 | all exons | 7 | |

**Table 4.1: Genomic regions interrogated by the initial panel and their mutation frequency in localised prostate cancer.** Interrogation of the gene set using cBioPortal demonstrates the frequency of SNVs reported in previous analyses of localised prostate cancer (Cerami *et al.;* Gao *et al.*). The formatting of this table is similar to that of **Table 3.1**.

#### 4.4.1.2 Prostate Specific Primer Panel

To allow a more comprehensive analysis, primers were designed to target genomic regions of interest in prostate and bladder cancer. Regions of interest were included based on the following factors:

- prevalence of mutation
- number of primers required to target region
- ease of targeting region (e.g. GC neutrality, avoiding repetitive regions, etc.)

Using the COSMIC database (Forbes *et al.*, 2011) and the results of major sequencing projects (e.g. Barbieri *et al.*, 2012; Grasso *et al.*, 2012; Guo *et al.*) I compiled a list of mutations and their prevalence in localised prostate cancer. This list was limited to include mutations that could be targeted with a manageable number of primer pairs (48-150 total), whilst allowing the broadest possible coverage of the patient population. Subsequently,

primers were designed to be as close to the ideal specifications as possible, set out in Table 2.3. Assays covering several regions of interest had previously been designed by other members of the Rosenfeld group and these were utilised with their permission for this study.

Primers were designed using Primer 3 (Untergasser *et al.*, 2012) and an automated pipeline developed by Franceso Marass (described in section 2.4.3). Through the above methods, a multiplexed prostate specific primer panel was designed targeting hotspot SNVs in *PIK3CA*, *HRAS*, *KRAS*, *PTEN*, *AKT1*, *NRAS*, *BRAF*, *EGFR*, *CTNNB1*, *IDH1* and entire exons of *SPOP*, *FOXA1*, *TP53*, *OR5L1* and *AR*. Table 4.2 lists the genes targeted in the multiplexed primer panel and illustrates an expected SNV detection rate of 30% for localised prostate cancer (Cerami *et al.*, 2012; Gao *et al.*, 2013; Cancer Genome Atlas Research, 2015). In total 150 primers were included in the final prostate specific primer panel, including assays from previous designs and those designed specifically for this project. Newly designed primer pairs were validated by BLAT searches (Kent *et al.*), insilico PCR (Kent *et al.*) and experimental validation by PCR with male genomic DNA (Promega, PR-G1471), as described in section 2.4.5.

Primer multiplexing was used on the final primer panel to overcome the limitation imposed by the 48 assay input channels in the Access Array microfluidic chip. Multiplexing of primers would ideally take into account primer pair annealing temperature, GC content, length and genomic location. However, due to the stringent requirements placed upon primer pairs during the design stage, primer pairs were considered as equivocal in these parameters and multiplexed pools were based primarily on genomic location (e.g. to avoid ‘nested’ primer pairs). The final multiplex pattern is shown in Figure 4.1.

**Previously Designed Primers**

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| A | BRAF_D0008_001F | EGFR_D0008_006F | EGFR_D0008_014F | KRAS_D0009_002F | PIK3CA_D0012_004F | TP53_D0008_005F | TP53_D0008_013F |
| B | BRAF_D0009_001F | EGFR_D0008_007F | EGFR_D0008_015F | KRAS_D0008_003F | PIK3CA_D0008_001F | TP53_D0008_006F | TP53_D0008_014F |
| C | BRAF_D0009_002F | EGFR_D0008_008F | EGFR_D0008_016F | NRAS_D0012_001F | PIK3CA_D0012_005F | TP53_D0008_007F | TP53_D0008_015F |
| D | EGFR_D0008_001F | EGFR_D0008_009F | EGFR_D0008_017F | NRAS_D0008_001F | PIK3CA_D0008_002F | TP53_D0008_008F | TP53_D0008_016F |
| E | EGFR_D0008_002F | EGFR_D0008_010F | EGFR_D0008_018F | NRAS_D0008_002F | TP53_D0008_001F | TP53_D0008_009F | TP53_D0008_017F |
| F | EGFR_D0008_003F | EGFR_D0008_011F | KRAS_D0008_001F | PIK3CA_D0012_002F | TP53_D0008_002F | TP53_D0008_010F | TP53_D0008_018F |
| G | EGFR_D0008_004F | EGFR_D0008_012F | KRAS_D0009_001F | PIK3CA_D0012_001F | TP53_D0008_003F | TP53_D0008_011F | TP53_D0008_019F |
| H | EGFR_D0008_005F | EGFR_D0008_013F | KRAS_D0008_002F | PIK3CA_D0012_003F | TP53_D0008_004F | TP53_D0008_012F | TP53_D0008_020F |

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | PTEN_E00001156315_1mF | PTEN_E00001156330_1F | | | | |
| B | PTEN_E00001156315_5F | PTEN_E00001156330_3F | | | | |
| C | PTEN_E00001156315_7F | PTEN_E00001156337_4F | | | | |
| D | PTEN_E00001156321_1F | PTEN_E00001156344_1F | | | | |
| E | PTEN_E00001156321_2F | PTEN_E00001156351_1F | | | | |
| F | PTEN_E00001156321_4mF | PTEN_E00001456541_1F | | | | |
| G | PTEN_E00001156327_1F | PTEN_E00001456541_2F | | | | |
| H | PTEN_E00001156327_4F | PTEN_E00001456562_1F | | | | |

**Newly Desigen Primers**

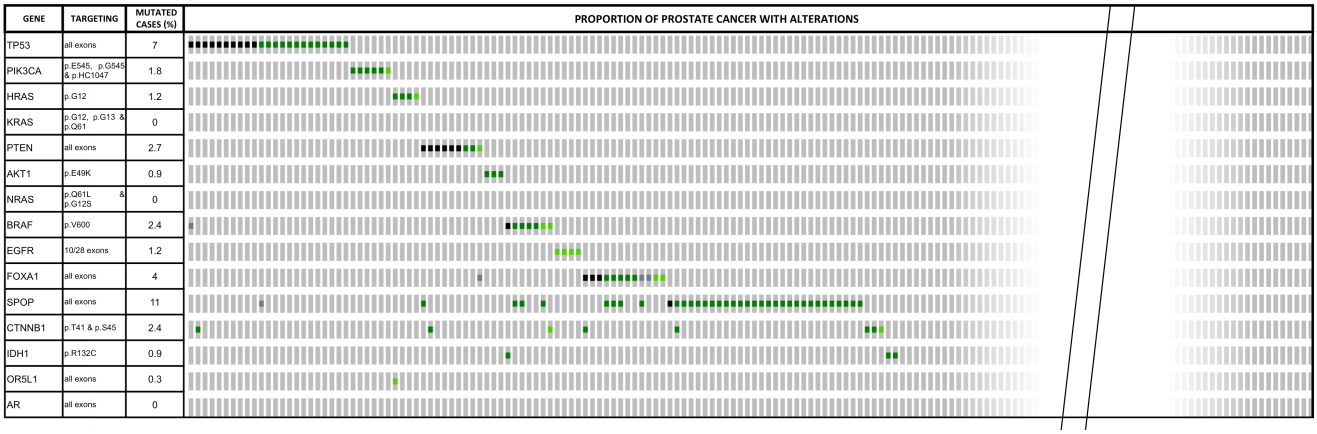
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| A | FOXA1_D0006_001F | FOXA1_D0006_009F | SPOP_D0006_005F | SPOP_D0006_013F | OR5L1_D0006_006F | AR_D0006_007F | AR_D0006_015F |
| B | FOXA1_D0006_002F | FOXA1_D0006_010F | SPOP_D0006_006F | SPOP_D0006_014F | OR5L1_D0006_007F | AR_D0006_008F | AR_D0006_016F |
| C | FOXA1_D0006_003F | FOXA1_D0006_011F | SPOP_D0006_007F | SPOP_D0006_015F | AR_D0006_001F | AR_D0006_009F | AR_D0006_017F |
| D | FOXA1_D0006_004F | FOXA1_D0006_012F | SPOP_D0006_008F | OR5L1_D0006_001F | AR_D0006_002F | AR_D0006_010F | AR_D0006_018F |
| E | FOXA1_D0006_005F | SPOP_D0006_001F | SPOP_D0006_009F | OR5L1_D0006_002F | AR_D0006_003F | AR_D0006_011F | AR_D0006_019F |
| F | FOXA1_D0006_006F | SPOP_D0006_002F | SPOP_D0006_010F | OR5L1_D0006_003F | AR_D0006_004F | AR_D0006_012F | AR_D0006_020F |
| G | FOXA1_D0006_007F | SPOP_D0006_003F | SPOP_D0006_011F | OR5L1_D0006_004F | AR_D0006_005F | AR_D0006_013F | AR_D0006_021F |
| H | FOXA1_D0006_008F | SPOP_D0006_004F | SPOP_D0006_012F | OR5L1_D0006_005F | AR_D0006_006F | AR_D0006_014F | AR_D0006_022F |

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| A | SPOP_D0020_001F | FOXA1_D0020_009F | | | | | |
| B | SPOP_D0020_002F | | | | | | |
| C | SPOP_D0020_003F | | | | | | |
| D | SPOP_D0020_004F | | | | | | |
| E | SPOP_D0020_005F | | | | | | |
| F | SPOP_D0020_006F | | | | | | |
| G | FOXA1_D0020_007F | | | | | | |
| H | FOXA1_D0020_008F | | | | | | |

**Multiplexed Plate**

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | | | | | | |
| B | | | | | | |
| C | | | | | | |
| D | | | | | | |
| E | | | | | | |
| F | | | | | | |
| G | | | | | | |
| H | | | | | | |

**Figure 4.1: Multiplexed primer pair 96 well template**. Primer pairs were multiplexed according to their colour code on this template. Care was taken to ensure the primer pairs with the same colour code were >5Kb from other primer pairs in the same well.



| GENE | TARGETING | MUTATED CASES (%) | PROPORTION OF PROSTATE CANCER WITH ALTERATIONS |
|---|---|---|---|
| TP53 | all exons | 7 | |
| PIK3CA | p.E545, p.G545 & p.HC1047 | 1.8 | |
| HRAS | p.G12 | 1.2 | |
| KRAS | p.G12, p.G13 & p.Q61 | 0 | |
| PTEN | all exons | 2.7 | |
| AKT1 | p.E49K | 0.9 | |
| NRAS | p.Q61L & p.G12S | 0 | |
| BRAF | p.V600 | 2.4 | |
| EGFR | 10/28 exons | 1.2 | |
| FOXA1 | all exons | 4 | |
| SPOP | all exons | 11 | |
| CTNNB1 | p.T41 & p.S45 | 2.4 | |
| IDH1 | p.R132C | 0.9 | |
| OR5L1 | all exons | 0.3 | |
| AR | all exons | 0 | |

**Table 4.2: Genomic regions interrogated by multiplex prostate specific panel in localised prostate cancer.** Analysis of the gene set by cBioPortal demonstrates the expected frequency of SNVs, based on the results of previous studies (Cerami *et al.;* Gao *et al.*). This table is formatted according to **Table 3.1**.

### 4.4.2 Exploring the utility of ctDNA analysis in aggressive Prostate Cancer

To explore whether ctDNA analysis be an informative biomarker in men with aggressive disease, in particular those who later go on to develop metastatic prostate cancer, I interrogated longitudinal samples collected at early and late disease stages. Following informed consent and ethical approval from the Royal Melbourne Hospital (Austrialia) Human Ethics committee, samples were collected from 19 patients with overt metastatic prostate cancer in collaboration with Mr. Mathew Hong and Dr. Chris Hovens. These samples from patients who went on to develop metastasis were compared to 19 patients with localised disease, which were collected at Addenbrooke's Hospital, see Section 4.4.1.

DNA was extracted from fresh frozen tissue and FFPE tissue using the QIAamp DNeasy Blood & Tissue Kit (69504) and QIAamp DNA FFPE Tissue kit (Qiagen 56404) according to the manufacturer's protocols (with modifications as outlined in the Rosenfeld Group SOPs). DNA from 1ml of plasma and 200µl of whole-blood samples were extracted using QIAamp circulating nucleic acid and QIAmp mini blood kits (Qiagen), respectively. To interrogate entire exonic regions of *TP53* alone, a specific amplicon panel was created from primers previously designed by Dr Tim Forshew. Details are listed in Appendix A-6. Samples from all 38 patients were analysed by performing tagged-amplicon deep sequencing (TAm-Seq). Sequencing libraries were prepared as previously described (Forshew *et al.*, 2012). Libraries were quantified using KAPA qPCR library quantification kits, and sequenced on the Illumina MiSeq platform using 150bp paired-end sequencing protocol over two lanes. Sequencing reads were demultiplexed according to sample-specific barcodes and aligned to the reference genome (hg19) using BWA (0.7.5a) (Li and

Durbin, 2009). Mutations were called and quantified as previously described (Forshew *et al.*, 2012).

### 4.4.3 Accounting for prostate heterogeneity in localised disease

To assess the representation of intratumoural heterogeneity in ctDNA, I secured collaboration with the International Cancer Genome Consortium (ICGC) to identify men whose prostatectomy specimens had undergone multi-region whole-genome sequencing (Cooper *et al.*, 2015). Only 2 of these men (case 7 and case 8) had pre-operative plasma blood draws. Plasma samples and multi-region whole-genome sequencing data to a median depth of 50x were acquired for these 2 men. DNA from the plasma samples were extracted using the Circulating Nucleic acid kit (QIAamp 55114) and quantified with dPCR as described in section 2.4.2.

Through collaboration with the ICGC Working Group details of mutations from each region of the tumours (SNVs and re-arrangements) and remaining tumour DNA were obtained. Figure 4.2 details the spatial relationships and Gleason grading of the regions analysed by whole-genome sequencing by the ICGC group.

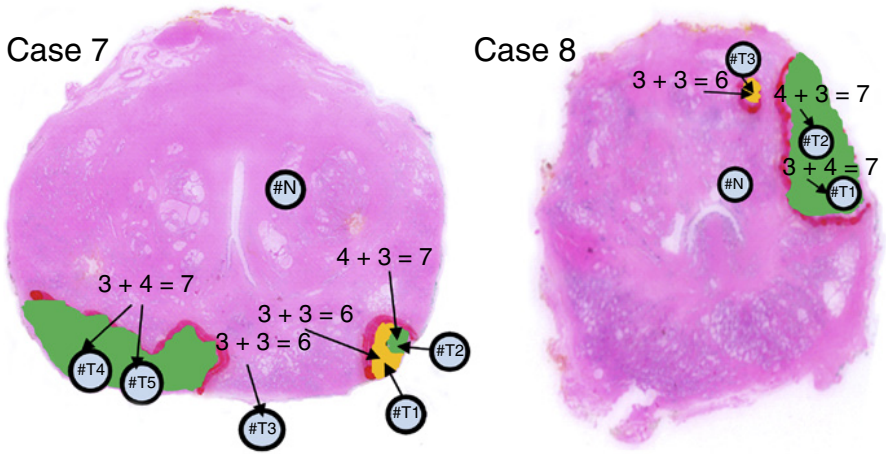



**Figure 4.2: Prostate samples sent for genome wide sequencing as part of ICGC prostate cancer project.** "*ERG* rearrangements determined by FISH. Case 7 had localised multifocal prostate cancer containing two separate foci (T1/T2/ T4/T5 and T3). Case 8 was also multifocal, with nodules T1/T2 and T3). Yellow, un-rearranged normal ERG gene; red, ERG gene split but both 3′ and 5′ ends retained; green, ERG gene rearranged but only its 3′ end retained (Del)" (Cooper *et al.*, 2015)**.** The differing ERG rearrangement status of the clones, suggest that regions may have unique mutational landscapes. (Figure and caption modified from Cooper *et al.,* 2015)

Patient- and clone-specific primers were designed to target cfDNA. Targets were selected based on their high AF in tumour samples and differences between inferred clones, including exonic, intronic, intergenic, mitochondrial SNVs and re-arrangement breakpoints.

Primers were validated in singleplex PCR reactions, adhering closely to TAm-Seq conditions. PCR products were screened for the expected product on 2% agarose e-Gels (Life technologies).

DNA was extracted from 1ml of pre-prostatectomy plasma, using the QIAamp Circulating Nucleic Acid kit (Qiagen). TAm-Seq was applied to the remaining DNA (extracted from each region for case 7 and 8 by the ICGC working group) and plasma DNA with the patient specific primers, as described previously (section 2.4.5., (Forshew *et al.*)).

## 4.5 Results

### 4.5.1 Few mutations are detectable in localised prostate cancer using a standard gene panel

For a proof of principle analysis, DNA was extracted from FFPE slides for 12 prostatectomy specimens. All 12 men had locally advanced prostate cancer and had no evidence of recurrence, Figure 4.3 and Table 4.3.

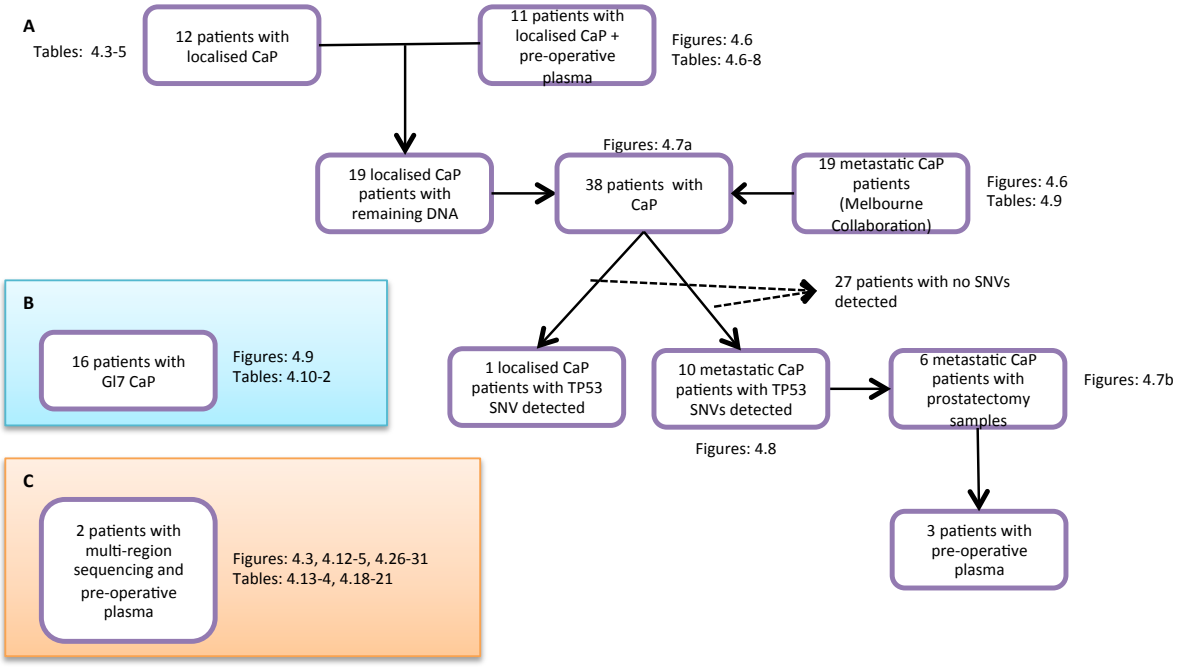



**Figure 4.3: Flow diagram of patient material used to investigate ctDNA in localised prostate cancer. A. Samples used for proof of principle ctDNA analysis in localised prostate cancer.** Samples from 12 and 11 men with localised prostate cancer were used to investigate the feasibility of ctDNA detection using a generic and prostate specific primer panel. Through collaboration with Melbourne University Hospital, samples from 19 men with metastatic prostate cancer were screened for exonic *TP53* SNVs and compared to a cohort of 19 men with localised disease. The reported prevalence of *TP53* mutations was 7% (Forbes *et al.*) and therefore, >14 men were required for this analysis. Men with localised disease were selected due to quantity of remaining DNA following proof of principle analysis. **B. Samples used to assess *TP53* SNVs in tissue samples. C. Sample flow to investigate prostate heterogeneity in localised disease**

| ID | Diagnosis PSA | Treatment option | Stage | Gleason | Last PSA value |
|---|---|---|---|---|---|
| KP1* | | | | | |
| KP2 | 6.3 | RRP | pT2c | 7 | 0.05 |
| KP3 | 9.5 | RRP + LN | pT3a | 7 | <0.1 |
| KP4 | 6.3 | RRP | pT2c | 6 | <0.1 |
| KP5 | 8 | RRP | pT2c | 7 | 0.02 |
| KP6 | 4.1 | RRP | pT2c | 7 | 0.01 |
| KP7 | 10.3 | RRP | pT3a | 7 | 0.02 |
| KP8 | 8.8 | RRP | pT3a | 7 | 0.02 |
| KP9 | 8.7 | RRP | pT3a | 7 | 0.02 |
| KP10 | 7 | RRP | pT3a | 7 | 0.02 |
| KP11 | 6.8 | RRP | pT2c | 7 | 0.11 |
| KP12 | 9.6 | RRP | pT2c | 9 | 0.02 |

**Table 4.3: Clinical characteristics of samples used in the proof-of-principle cohort.** *Clinical details for case KP1 were unavailable.

Table 4.4 shows the concentration of DNA extracted from the FFPE slides. Fifty nanograms of extracted DNA used as input for the TAm-Seq protocol. MiSeq 150bp paired end sequencing resulted in 8.5 million reads of which 7 million aligned to the reference genome. In total 3 SNVs were found in 2/12 of men, all of which were exonic and non-synonymous (see Table 4.5). These results demonstrate the technical feasibility of TAm-Seq to detect SNVs in tissue DNA from men with locally advanced prostate cancer. The low detection rate is likely due to the intra-and inter-tumour heterogeneity of localised prostate cancer, and hence an extended primer panel would be required to increase the number of men in which mutations are detected.

| Sample | DNA concentration (ng/µl) |
|---|---|
| KP1 | 11.6 |
| KP2 | 31 |
| KP3 | 20.2 |
| KP4 | 36.1 |
| KP5 | 23.3 |
| KP6 | 10.2 |
| KP7 | 141 |
| KP8 | 38.5 |
| KP9 | 91.9 |
| KP10 | 59.9 |
| KP11 | 40.9 |
| KP12 | 33.2 |

**Table 4.4: Qubit fluorometer DNA quantification of DNA extracted from FFPE embedded tissue for the pilot prostate TAm-Seq analysis.**

| Sample ID | AF | Mutation (hg19) | Effect |
|---|---|---|---|
| KP6 | 0.0545 | chr10:89720683_C>G | PTEN, p.F278L |
| KP6 | 0.031 | chr7:55259530_G>A | EGFR, p.G863D |
| KP9 | 0.139 | chr7:55259452_A>C | EGFR, p.D837A |

**Table 4.5: Single Nucleotide variants in two samples from patients with localised prostate cancer.**

### 4.5.2 Using multiple targets to improve detection in localised prostate cancer

To improve the yield of mutations detected by TAm-Seq in localised prostate cancer, I developed an expanded prostate panel.

#### 4.5.2.1 Prostate Specific Panel validation

Primer specificity was determined by PCR using control DNA extracted from the white blood cells of healthy male volunteers (MCDNA). PCR conditions were similar to that which would occur in TAm-Seq. PCR conditions were designed to match the cycling

conditions used in TAm-Seq. Primer specificity was assessed by checking that PCR products were of the expected length.

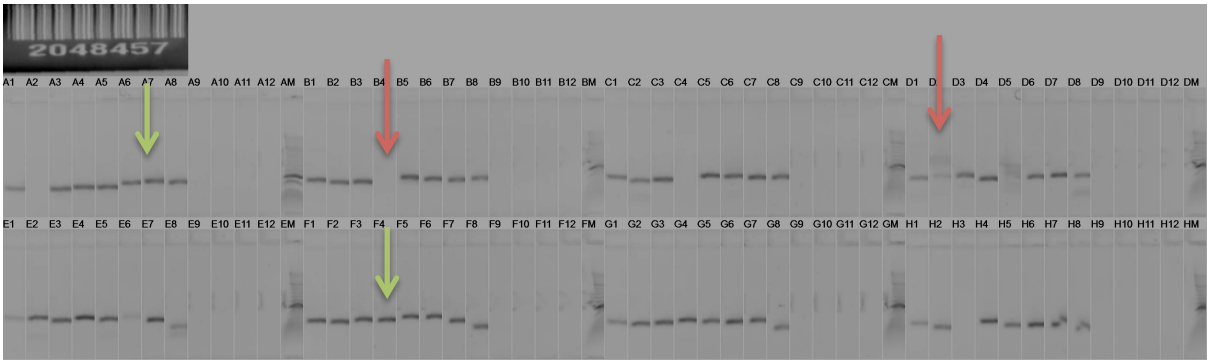



**Figure 4.4: Agarose eGel (Invitrogen) image following PCR amplification of 65 newly designed primers for the prostate cancer panel.** The red arrows show amplicons that have failed QC and the green arrows highlight examples where DNA size bands are within the expected range for the PCR products.

Figure 4.4 shows an image of an eGel following PCR amplification of 65 new prostate primer pairs including targets for *FOXA1, SPOP, OR5L1, AR, AKT1, HRAS* & *IDH1*. Some amplicons failed or produced non-specific products, and where possible were redesigned. Figure 4.5 shows 3 primer pairs that successfully amplified target loci following redesign. Of the 64 new primer pairs that were designed, 5 regions failed QC despite redesigning. These 5 loci represented infrequently mutated regions of *SPOP* and *FOXA1* and were a small fraction of the SNVs detected in localised prostate cancers. Therefore, the remaining 59 new primer pairs were used to generate a final 149 loci prostate cancer panel that was used for further analysis. The final primer panel is detailed in Appendix A-7.


**Figure 4.5: Agarose eGel (Invitrogen) image analysis following PCR amplification of duplicate results of 3 re-designed primer pairs with positive (lane 1-6) and negative (lane 10) controls.**

### 4.5.3 Localised prostate tumour tissue contains few mutations detected by prostate specific TAm-Seq

The prostate specific primer panel was used to perform TAm-Seq on FFPE tissue samples from 11 men with locally advance prostate cancer with no evidence of recurrence on latest PSA analysis, with median follow up of 54 months, see Table 4.6.

| Patient ID | Diagnosis PSA | treatment option | Stage | Gleason | Last PSA value | Follow up period (months) |
|---|---|---|---|---|---|---|
| Matched 5 | 2.86 | RRP | pT2C | 6 | 0.1 | 67 |
| Matched 6 | 11.6 | RRP | pT3a | 7 | 0.1 | 51 |
| Matched 11 | 5.3 | RRP | pT2C | 6 | 0.1 | 61 |
| Matched 3 | 4.6 | RRP | pT3a | 7 | 0.1 | 40 |
| Matched 4 | 6.9 | RRP | pT2c | 6 | 0.1 | 63 |
| Matched 1 | 9.6 | RRP | T2c | 7 | 0.1 | 56 |
| Matched 2 | 89 | RRP | T3b | 9 | 0.1 | 4 |
| Matched 10 | 11 | RRP | T2a | 7 | 0.1 | 42 |
| Matched 9 | 7.5 | RRP | T3a | 7 | 0.1 | 41 |
| Matched 8 | 8.3 | RRP | T2c | 7 | 0.1 | 9 |
| Matched 7 | 8.5 | RRP | T3a | 6 | 0.1 | 59 |

**Table 4.6: Clinical details of 11 men with localised prostate cancer subjected to TAm-Seq analysis using the prostate specific primer panel.**

Five patients had areas of normal prostate tissue, marked by an academic uropathologist (Dr. Anne Warren), on the same slide as the tumour. These “benign” areas were macro dissected and DNA extracted for use as matched normal controls, see Table 4.7. DNA extraction from FFPE prostatectomy samples yielded a median eluate concentration of 31.4 ng/μL (ranging 6.8 ng/μL - 247 ng/μL) as measured using DNA specific fluorescent dye (Qubit, Thermofisher), see Table 4.7.

| Sample name | Concentration ng/µL |
|---|---|
| Matched_1 | 29.8 |
| Matched_2 | 28.9 |
| Matched_3 | 81 |
| Matched_4 | 65.3 |
| Matched_4N | 26.3 |
| Matched_5 | 32.9 |
| Matched_5N | 4.91 |
| Matched_6 | 247 |
| Matched_6N | 24.4 |
| Matched_7 | 25.8 |
| Matched_8 | 6.83 |
| Matched_9 | 61 |
| Matched_9N | 18.6 |
| Matched_10 | 14.3 |
| Matched_10N | 49.2 |
| Matched_11 | 21.3 |

**Table 4.7: Qubit fluorometer DNA quantification of DNA extracted from prostate FFPE embedded tissue of 11 men and matched normal tissue (where applicable).**

Pooled TAm-Seq libraries were re-sequenced on an Illumina HiSeq 2000 using 125bp paired end chemistry and resulted in 74.6 million aligned reads. Despite this the median coverage was only 333x and 35.5% (53/149) of primers had a median coverage of 100x or less. These poor coverage primers targeted regions in *AR*, *FOXA1*, *SPOP*, *PTEN* and *TP53*, see Figure 4.6. Subsequently, only 1 SNV was detected in patient 11 from an intergenic region of *AR* (Table 4.8).

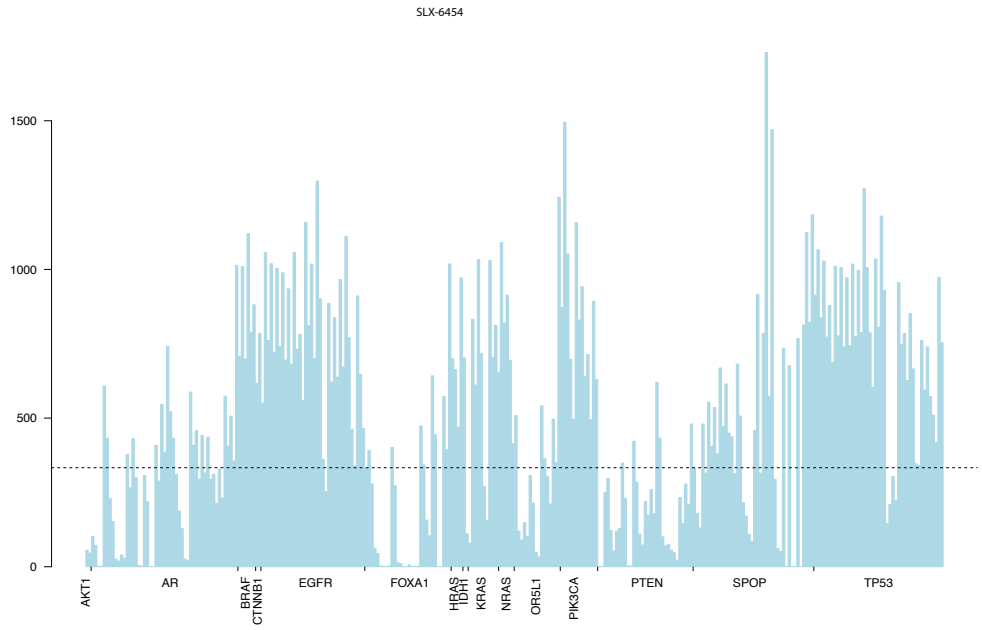



**Figure 4.6: Median amplicon coverage for TAm-Seq using the prostate primer panel.** Bars represent the median coverage (y axis) for each amplicon in the panel (x-axis). The dotted line represents the overall median coverage (333x).

| Patient ID | AF | Mutation (hg19) | Effect |
|---|---|---|---|
| Matched 7 | 0.1644 | Chr23:66765798_C>T | AR intergenic |

**Table 4.8: Single Nucleotide variant detected in one patient with Tam-Seq using the extended prostate panel.**

The AR intergenic SNV was not previously catalogued by the COSMIC database (Forbes *et al.*). Due to the intergenic nature is of uncertain tumourigenic effect. The low detection rate of SNVs in this experiment is likely to be multi-factorial. Firstly regions with low coverage will have reduced detection rates for SNVs at low AF in those regions. Secondly, the broad mutational spectrum of prostate cancer, with a long tail of infrequently mutated genes (Barbieri *et al.*), and a low mutation rate of 0.9/Mb (Berger *et al.*, 2011) may result in poor detection rate despite attempting to cover almost 22Kb of the genome. Furthermore, the utility of a clinical ctDNA test would be further obfuscated by prostate tumour heterogeneity and I subsequently explored this hypothesis. However, capture based sequencing will allow the interrogation of increased genomic regions and urinary or semen cfDNA may concentrate prostate specific cfDNA. As the cost of sequencing continues to drop (Hayden, 2014) high coverage capture sequencing of cfDNA from plasma, urine and/or semen may improve the utility of cfDNA analysis in localised prostate cancer.

### 4.5.4 Clones that dominate in metastatic castrate resistant prostate cancer are present at a localised stage

To evaluate whether ctDNA could be an informative biomarker of aggressive cancer, on samples were collected from 19 patients with metastatic disease and 19 with localised disease. The collection comprised of 91 samples, including 48 metastatic tissue, /blood/ and plasma samples taken at or after the time of clinical relapse and 43 primary tissue, /blood and /plasma samples taken at or before the time of prostatectomy.

Local samples were chosen from the pilot cohorts described above and were included base on the quantity of DNA remaining. Samples from metastatic patients yielded a mean of 26.0 ng/μL, 98.0 ng/μL, 3647.3 ng/μL and 6.4 ng/μL, for FFPE tissue, fresh frozen tissue, whole-blood and plasma, see Table 4.9.

| Sample ID on tube | Sample Type | DNA ng/μl |
|---|---|---|
| | Patient 002 | |
| 002_blood | whole-blood | 18.8 |
| 002_plasma | plasma | 6.35 |
| 002_tissue01 | tissue | 66 |
| 002_tissue02 | tissue | 59.6 |
| 002_tissue03 | FFPE tissue | 0.652 |
| 002_tissue04 | FFPE tissue | 1.01 |
| | Patient 014 | |
| 014_blood01 | whole-blood | 30.8 |
| 014_plasma | plasma | 10.37 |
| 014_tissue01 | tissue | 48.7 |
| 014_tissue02 | FFPE tissue | 63.6 |
| 014_tissue03 | FFPE tissue | 49 |
| 014_tissue04 | tissue | 49 |
| | Patient 067 | |
| 067_blood01 | whole-blood | 4922.5 |
| 067_blood02 | whole-blood | 20 |
| 067_tissue01 | tissue | 37.6 |
| 067_tissue02 | tissue | 70 |
| | Patient 068 | |
| 068_blood | whole-blood | 4922.5 |
| 068_Tissue02 | tissue | 62.4 |
| 068_Tissue03 | tissue | 91 |
| 068_Tissue04 | tissue | 29 |
| 068_Tissue05 | tissue | 32.8 |
| | Patient 094 | |
| 094_blood | whole-blood | 4922.5 |
| 094_tissue | tissue | 97.8 |
| | Patient 130 | |
| 130_blood | whole-blood | 4922.5 |
| 130_tissue | tissue | 106 |
| | Patient 152 | |
| 152_blood | whole-blood | 4922.5 |
| 152_tissue | tissue | 97 |
| | Patient 177 | |

| | | |
|---|---|---|
| 177_blood | whole-blood | 4922.5 |
| 177_blood02 | whole-blood | 30.8 |
| 177_tissue01 | tissue | 78.6 |
| 177_tissue02 | tissue | 1.94 |
| 177_tissue03 | tissue | 49 |
| 177_tissue04 | tissue | 41 |
| 177_tissue05 | tissue | 27.8 |
| | Patient 188 | |
| 188_plasma | plasma | 4.65 |
| 188_tissue | tissue | 86.4 |
| | Patient 197 | |
| 197_blood | whole-blood | 4922.5 |
| 197_tissue | tissue | 65 |
| | Patient 252 | |
| 252_blood01 | whole-blood | 4922.5 |
| 252_blood02 | whole-blood | 4922.5 |
| 252_Tissue | tissue | 96.6 |
| | Patient 298 | |
| 298_tissue01 | tissue | 65.8 |
| | Patient 299 | |
| 299_blood01 | whole-blood | 4922.5 |
| 299_plasma | plasma | 7.4 |
| 299_tissue01 | tissue | 91.8 |
| 299_tissue04 | FFPE tissue | 15.8 |
| 299blood_02 | whole-blood | 84.8 |
| 299tissue_02 | tissue | 66.2 |
| 299tissue_03 | tissue | 95.4 |
| | Patient 486 | |
| 486_blood | whole-blood | 4922.5 |
| 486_plasma | plasma | 6.35 |
| 486_tissue01 | tissue | 604 |
| 486_tissue02 | tissue | 102 |
| | Patient 498 | |
| 498_blood01 | whole-blood | 4922.5 |
| 498_blood02 | whole-blood | 19.9 |
| 498_plasma | plasma | 3.17 |
| 498_tissue02 | tissue | 66.6 |
| 498_tissue03 | tissue | 54 |
| 498_tissue04 | tissue | 1.59 |
| | Patient 513 | |
| 513_blood | whole-blood | 4922.5 |
| 513_tissue01 | tissue | 302 |
| 513_tissue02 | tissue | 66 |
| | Patient 650 | |
| 650_blood | whole-blood | 4922.5 |
| 650_tissue01 | tissue | 236 |
| 650_tissue02 | tissue | 49 |
| | Patient 752 | |
| 752_blood | whole-blood | 4922.5 |
| 752_tissue | tissue | 3.54 |
| | Patient 854 | |
| 854_blood | whole-blood | 4922.5 |
| 854_tissue | tissue | 130 |
| | Patient 932 | |
| 932_plasma | whole-blood | 4922.5 |
| 932_tissue01 | tissue | 400 |
| 932_tissue02 | tissue | 96.2 |

**Table 4.9: DNA yield of samples obtained from men with metastatic prostate cancer.** Tissue = fresh frozen tissue.

TAm-Seq was initially performed on DNA extracted from each sample and was sequenced to an average depth of 7,848x permitting detection of low-frequency *TP53* mutations. The

results (summarised in Figure 4.7, Figure 4.8 and Appendix A-8) show that 10 out of 19 metastatic patients had detectable *TP53* mutations, compared with one out of the 19 patients with localisedlocalized disease. High depth targeted sequencing showed that *TP53* mutations were detectable in primary tumours from men who went on to develop metastatic disease, in six of the 10 metastatic patients with mutant *TP53* for which a matching primary sample was available (Figure 4.7b). Furthermore, in most of these cases, we were able to detect these mutations in the blood or plasma.

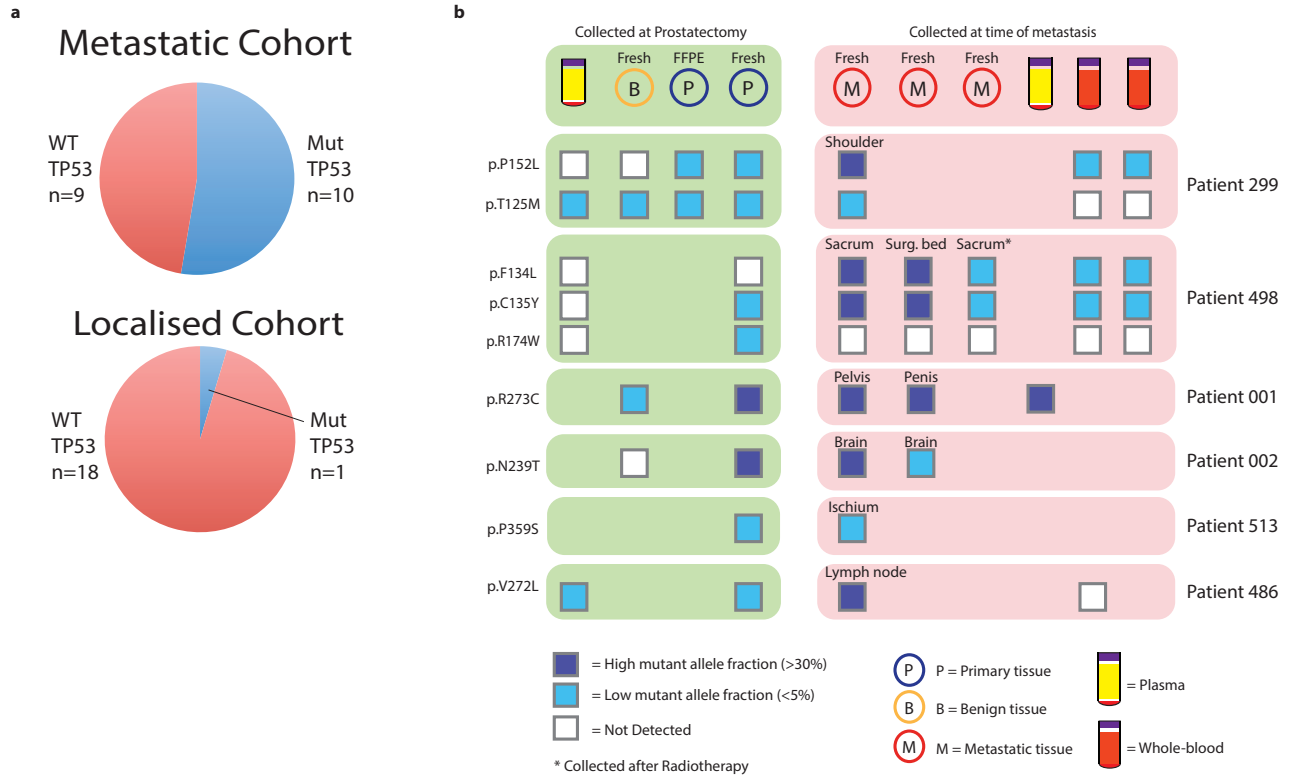



**Figure 4.7: *TP53* mutations identified via TAm-Seq. (a)** Pie charts representing the number of patients with detected *TP53* mutations across the metastatic and localised cohorts. **(b)** A schematic indicating the presence of *TP53* mutations and their allele frequency for all patients in the metastatic cohort that had matched primary and metastatic samples.

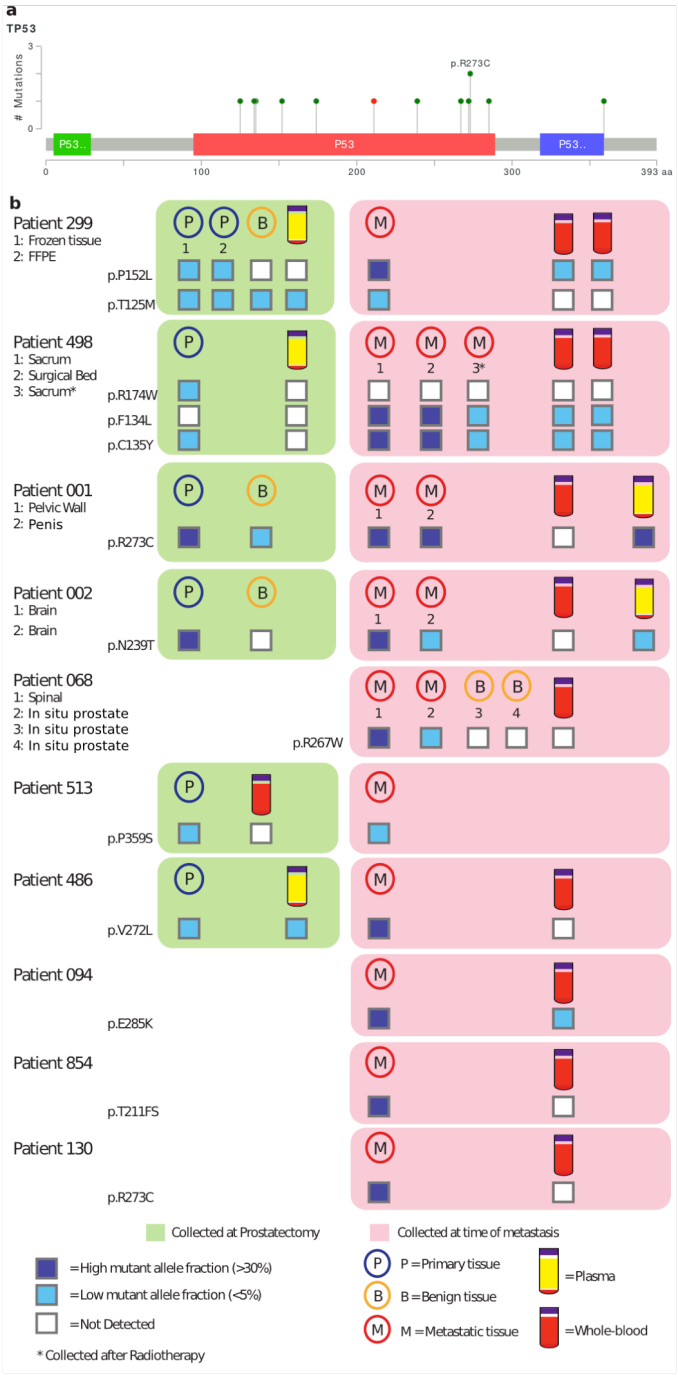



**Figure 4.8: *TP53* mutations identified via TAm-seq.** (**a**) A schematic indicating the location of each of the identified *TP53* mutations within the protein. Mis-sense mutations are shown in green and frame-shift in red. b) The detection results for the application of targeted amplicon sequencing of *TP53* across 72 samples. Only patients with a *TP53* mutation are listed (n=10), the remaining nine patients who were negative for *TP53* mutations are not shown.

This enrichment of *TP53* mutations in the metastasis samples, which appear at low frequency in primary samples, strongly suggests that the acquisition of *TP53* mutations in the prostate cancer setting enhances the metastatic potential of tumour subclones. Further supporting this observation is the additional *TP53* mutations found in patients 299 and 498. In both patients, a *TP53* mutation was detected at high frequency in the metastases, and this mutation was detected at low frequency in the primary. However, a second *TP53* mutation was detected at low frequency in the primary tumour, which was not found at high frequency in the metastasis. If these low frequency, secondary *TP53* mutations in the primary tumour confer metastatic potential, then the clones containing this mutation should be more likely to be detected in the blood/plasma of these patients, as was observed. The functional consequences of these secondary *TP53* mutations, as well as the cellular contexts in which they occurred, should also be taken into account when assessing the likely metastatic potential of these clones. To achieve this, future studies using single-cell sequencing and phenotypic validation will be required.

### 4.5.5 *TP53* SNVs in tissue samples are unlikely to accurately predict biochemical recurrence

The relatively high incidence of *TP53* mutations in metastatic samples has previously been described (Robinson *et al.*, 2015; Grasso *et al.*, 2012; Barbieri *et al.*, 2012). The data described above (Section 4.5.4) suggests that these potentially malignant clones can be detected directly from primary tumour tissue taken at the time of prostatectomy. To explore the potential utility of sequencing *TP53* in tumour tissue to predict aggressive disease, I investigated the prevalence of *TP53* mutations in a cohort of 16 men with Gleason 7 localised prostate cancer. Eight men developed biochemical recurrence within 5 years whilst 8 did not. Clinical characteristics are provided in Table 4.10.

| Pt ID | Age at diagnosis | PSA at diagnosis | Gleason Grade | Treatment | Treatment date | Pathological Grade | LN | Last PSA | Last PSA date | Biochemical recurrence |
|---|---|---|---|---|---|---|---|---|---|---|
| RELP_1 | 69 | 5.4 | 3+4=7 | RALP | 25/11/10 | pT2c | 0/5 | <0.02 | 01/12/15 | N |
| RELP_2 | 58 | 4.2 | 4+3=7 | RALP | 01/05/13 | pT2c | 0 | <0.02 | 04/08/16 | N |
| RELP_3 | 71 | 12.2 | 4+3=7 | RALP | 01/12/11 | pT2a | x | <0.05 | 19/08/16 | N |
| RELP_4 | 60 | 13.72 | 3+4=7 | RALP | 13/06/12 | pT2c | 0/1 | <0.1 | 01/11/16 | N |
| RELP_5 | 61 | 0 | 3+4=7 | RALP | 01/09/12 | pT2c | 0/10 | <0.02 | 13/12/16 | N |
| RELP_6 | 65 | 3.53 | 3+4=7 | RALP | 29/11/12 | pT2c | x | <0.1 | 24/10/16 | N |
| RELP_7 | 55 | 1.8 | 3+4=7 | RALP | 10/04/13 | pT2a | x | <0.02 | 10/02/14 | N |
| RELP_8 | 66 | 14.58 | 3+3=7 | RALP | 23/07/13 | pT2c |  | <0.02 | 07/06/16 | N |
| RELP_9 | 58 | 4.3 | 3+4=7 | RALP | 21/04/08 | pT2a | 0 | 0.18 | 26/09/16 | Y |
| RELP_10 | 60 | 5.8 | 3+4=7 | RALP | 12/11/08 | pT3a | x | <0.1 | 31/01/11 | Y |
| RELP_11 | 63 | 7.7 | 4+3=7 | RALP | 07/01/09 | T1c | N0 | 2.2 | 25/10/16 | Y |
| RELP_12 | 52 | 8.4 | 3+4=7 | RALP | 07/05/09 | pT2c | N0 | 0.31 | 28/10/14 | Y |
| RELP_13 | 66 | 11.7 | 3+4=7 | RALP | 26/04/10 | pT3a | pN1 | 0.43 | 15/10/10 | Y |
| RELP_14 | 58 | 4.85 | 3+4=7 | RALP | 09/09/10 | pT2c | pN0 | 0.98 | 11/11/15 | Y |
| RELP_15 | 65 | 6.2 | 4+3=7 | RALP | 11/02/12 | pT3a | pN0 | 0.26 | 13/12/16 | Y |
| RELP_16 | 66 | 5.07 | 3+4=7 | RALP | 10/10/12 | pT2c | pN0 | <0.02 | 05/09/16 | Y |

**Table 4.10: Clinical characteristics of men with localised prostate cancer used to determine the prognostic utility of *TP53* mutation analysis.** Sixteen men were identified retrospectively from the PROmpT study, with 8 men remaining recurrence free for 4 years post surgery (range: 0-7 years). The other 8 men developed biochemical recurrence within 4 years (range: 0-8 years), 2 of these men remain in biochemical remission following adjuvant therapy at the census date. RALP – Robotic Assisted Laparoscopic Surgery.

FFPE cores were generated from the dominant tumour foci and from adjacent benign regions of prostatectomy specimens and DNA was extracted with DNA-FFPE-Tissue (QIAmp, 56404), with paraffin removal performed by the heptane methods as described in the manufacturers instructions in 50µL of eluate. A mean concentration of 92.3 ng/µL of DNA was extracted from the samples (Table 4.11).

| Sample ID | Tissue Type | DNA concentration (ng/µL) |
|---|---|---|
| 1_B | BENIGN | 60 |
| 2_B | BENIGN | 142 |
| 3_B | BENIGN | 128 |
| 4_B | BENIGN | 135 |
| 5_B | BENIGN | 95.4 |
| 6_B | BENIGN | 51 |
| 7_B | BENIGN | 73.4 |
| 8_B | BENIGN | 114 |
| 9_B | BENIGN | 91.6 |
| 10_B | BENIGN | 130 |
| 11_B | BENIGN | 26 |
| 12_B | BENIGN | 70.8 |
| 13_B | BENIGN | 226 |
| 14_B | BENIGN | 62 |
| 15_B | BENIGN | 161 |
| 16_B | BENIGN | 88.8 |
| 1_T | TUMOUR | 49.6 |
| 2_T | TUMOUR | 90.6 |
| 3_T | TUMOUR | 101 |
| 4_T | TUMOUR | 122 |
| 5_T | TUMOUR | 148 |
| 6_T | TUMOUR | 8.96 |
| 7_T | TUMOUR | 64.4 |
| 8_T | TUMOUR | 123 |
| 9_T | TUMOUR | 109 |
| 10_T | TUMOUR | 28.8 |
| 11_T | TUMOUR | 57.8 |
| 12_T | TUMOUR | 73.4 |
| 13_T | TUMOUR | 146 |
| 14_T | TUMOUR | 26 |
| 15_T | TUMOUR | 79.4 |
| 16_T | TUMOUR | 71.2 |

**Table 4.11: DNA yield from FFPE prostatectomy slides in 16 men treated for Gleason 7 localised prostate cancer.**

Fifty nanograms of DNA samples were used as input for the TAm-Seq workflow using the *TP53* primer panel, as described in section 4.4.2 and detailed in Appendix A-6. Libraries were pooled and quantified before sequencing on a MiSeq (Illumina) with 150bp paired-end chemistry. Figure 4.9 shows the median coverage for all amplicons with a median depth of over 10,000x from the 27 million reads.

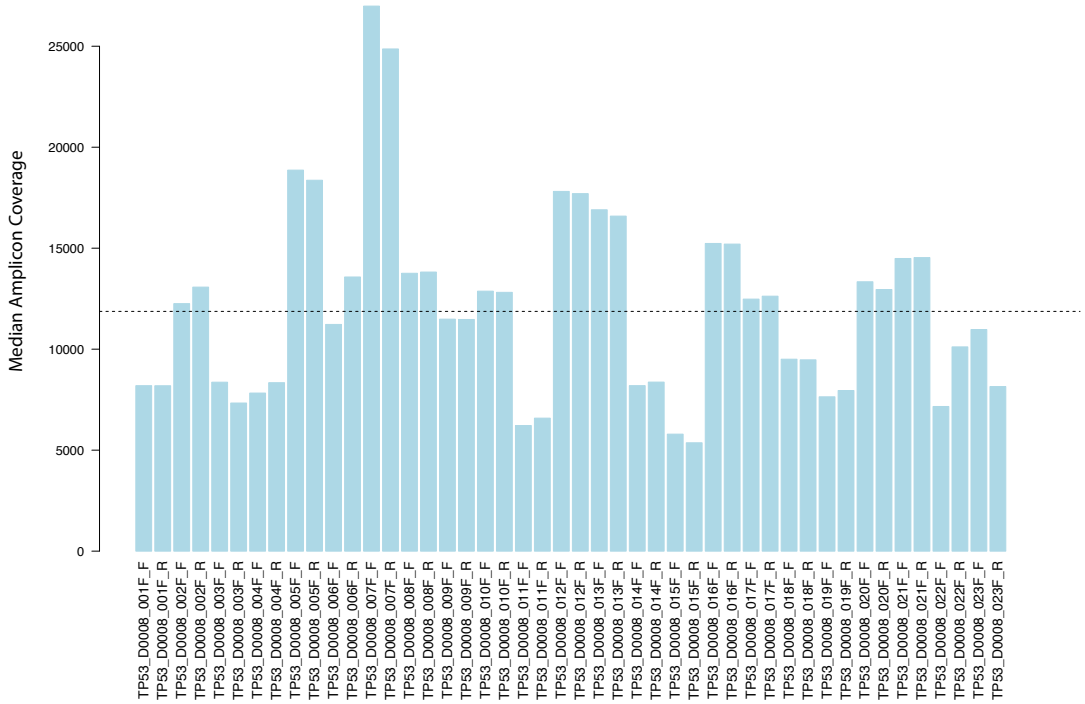



**Figure 4.9: Median coverage for amplicons in the *TP53* panel, showing each amplicon (x axis) vs. the median depth of sequencing coverage (y axis) for all samples.** Dotted line represents overall median coverage.

SNVs in *TP53* were detected in 3 patients, 2/8 patients with *TP53* SNVs did not have biochemical recurrence whilst only 1/8 men with *TP53* SNVs did go on to develop biochemical recurrence, see Table 4.12. Although the cohort of patients in this exploratory study was small, it seems that *TP53* analysis of individual prostatectomy specimens would not prove clinically useful due to the low sensitivity and specificity to predict biochemical recurrence. This may be due to the relatively short follow-up for this cohort and that the cohort was followed up for biochemical recurrence, which may not correlate to prostate cancer specific mortality (Freedland *et al.*, 2005; Pound *et al.*, 1999). However, only the dominant clone of the prostatectomy sample was analysed for *TP53* status in this study and, as prostate cancer is a heterogeneous cancer, the results of *TP53* analysis are affected if the clinically significant clone is not included in the analysed sample. Therefore, the use of biopsy material for *TP53* analysis would be problematic due to the well-documented concerns that standard needle biopsy sampling may miss a significant clone in prostate cancer (Eichler *et al.;* Patel and Tsui). For these reasons, the development of a tissue based approach for *TP53* analysis was not investigated further. However, ctDNA has been shown to represent multiple clones in metastatic cancer (Murtaza *et al.*, 2013). Therefore, *TP53* analysis of ctDNA could overcome the problems of prostate heterogeneity and is worthy of further investigation in a well-annotated cohort.

| Patient Id | Mutation (hg19) | Effect | AF |
|---|---|---|---|
| 13 | Chr17:7573983_C>T | TP53 p.L309L | 0.2 |
| 7 | Chr17:7578419_C>A | TP53 p.E132X | 0.001 |
| 2 | Chr17:7579472_G>A | TP53 p.P33L | 0.001 |

**Table 4.12: *TP53* SNVs detected in prostatectomy samples in men with Gleason 7 prostate cancer.**

### 4.5.6 Investigating prostate heterogeneity in localised disease

To investigate the ability of ctDNA to overcome prostate heterogeneity in patients with localised prostate cancer, patient and clone specific primers were designed and evaluated. Figure 4.10 depicts a 2% agarose e-Gels (Life technologies) of amplicon products following singleplex PCR with primer pairs designed to target clones identified from case 7. The majority of primer pairs produced a single PCR product in the expected size range. However, over both panels, 6 primer pairs did not produce a single band. As e-Gel analysis of PCR product may have a reduced sensitivity over re-sequencing and since the number of failed primers was low, the entire primer panel was used for further analysis.

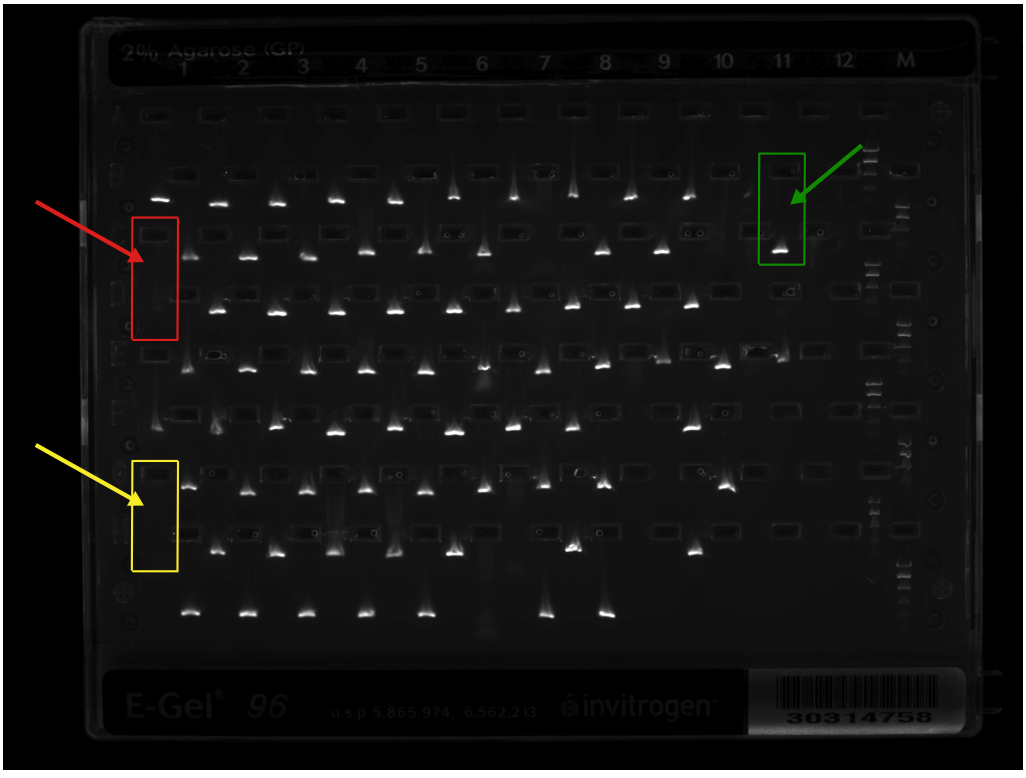



**Figure 4.10: Image of e-Gel PCR product analysis for primers designed to target case 7 clones.** Most primer pairs produced a single PCR product of the expected size (green arrow). Primer pairs targeting breakpoints did not produce any product due to the normal DNA template used in the reaction (yellow arrow). Six primers that were expected to produce a specific product did not (e.g. red arrow), although 2/6, had double product bands (both for case 8, data not shown).

A total of 77 and 41 primers were included for the analyses of case 7 and 8 respectively. Final primer panels are detailed in appendices A-9 and A-10. The 77 primers for Case 7 were multiplexed in to 48 wells, such that each well contained 1-3 primer pairs, see Figure 4.11. Whole Genome Amplification (WGA) (REPLI-G, Qiagen) was applied to 10ng DNA extracted from clonal regions and amplified DNA quantities listed in Table 4.13. TAm-Seq was performed using case 7 and 8 primer panels with WGA DNA template. PCR products were harvested, barcoded, pooled and quantified as previously described (Forshew *et al.*, 2012). Libraries for case 7 and case 8 were pooled and submitted for 125bp paired-end sequencing on an Illumina MiSeq.

| CRUK ID | Input DNA (ng) | Output DNA in ng/50µL |
|---|---|---|
| Case_7_N | 10 | 107 |
| Case_7_T1 | 10 | 103 |
| Case_7_T2 | 10 | 29.3 |
| Case_7_T4 | 10 | 65.5 |
| Case_7_T3 | 10 | 105 |
| Case_7_T5 | 10 | 122 |
| Case_8_T3 | 10 | 259 |
| Case_8_T1 | 10 | 72 |
| Case_8_T2 | 10 | 26.8 |
| Case_8_N | 10 | 44.1 |
| MCDNA | 10 | 54 |
| NTC | 0 | 5.1 |

**Table 4.13: Table of DNA amounts from multiregional patients with prostate cancer before and after whole genome amplification (REPLI-G, Qiagen).** N - adjacent normal tissue, T - clonal regions (as depicted in **Figure 4.2**). MCDNA – Male genomic DNA (Promega, PR-G1471), NTC – No Template Control.


**Figure 4.11: Multiplex planning for Case 7 primer pairs**. Primer pairs were multiplexed into 48 wells according to their colour code. The multiplexing plan was designed by a bioinformatic pipeline developed by Dr. James Morris (Senior Bioinformatics Analyst, Rosenfeld group), which accounted for the amplicon GC content, annealing temperature, and proximity of primers as detailed in section 4.4.1.2.

Sequencing resulted in 3.48 and 3.81 million reads aligning to the human genome from case 7 and case 8 barcoded samples. The median read depth was 790 and 1700 for case 7 and 8 respectively, analysis of reads aligning to each amplicon revealed that some amplicons had no coverage. Indeed, no primer pairs targeting re-arrangement breakpoints produced a PCR product that aligned to the expected region, neither did a further 2 and 3 primer pairs from case 7 and 8 respectively, see Figure 4.12. Seventy and 37 primer pairs successfully targeted regions of interest for case 7 and 8 respectively. There was no association between primer pair position and failure rate.

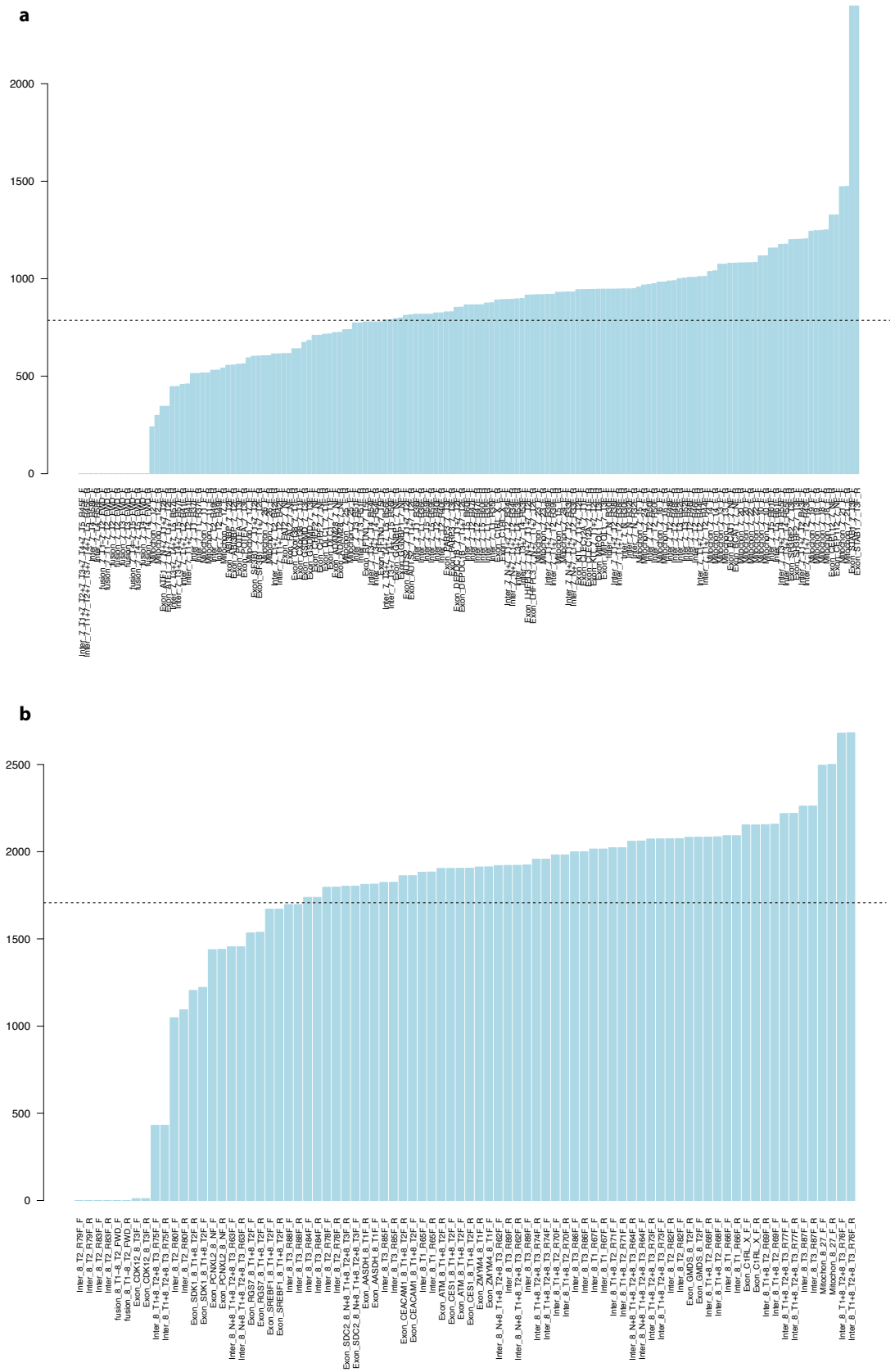



**Figure 4.12: Median coverage (y axis) per amplicon (x axis) for Case 7 (a) and Case 8 (b).**

None of the structural variant primers produced the expected product and to investigate whether this was caused by allelic drop out during whole genome amplification I performed singleplex PCR with fusion primers with unamplified DNA from the relevant tumour foci. Fusion PCR products were run on a DNA1000 bioanalyser chip (Agilent). Three primer pairs produced a specific product in the expected range, whilst one primer pair had a product in the expected range of primer dimer and the remaining primer pair generated non-specific products. The bam files from the WGA DNA experiment were interrogated for sequences covering the breakpoints using samtools (Li *et al.*, 2009) however, no fusion sequences were detected. The lack of fusion PCR product may be due to the combination of a single multiplexed PCR reaction followed by multiple singleplex PCR reactions during TAm-Seq. It is possible that primer pairs targeting DNA re-arrangements have far fewer template DNA molecules during the initial multiplexed PCR than the other primer pairs, which amplify wild type and ctDNA alike. Subsequently, little re-arranged DNA would be amplified resulting in reduced chance of pairing with re-arrangement primer pairs in the microfluidic singleplex PCR chamber. Fusion amplicons may work in direct singleplex PCR reactions, but due to the need for a separate reaction, template DNA was spared for multiplex TAm-Seq and so re-arrangement primer pairs were not included for further analysis.

#### 4.5.6.1 Heterogeneity in localised prostate cancer is not well represented in plasma ctDNA using standard TAm-Seq

Levels of ctDNA in localised prostate cancer were assessed using the patient specific primers detailed in section 4.4.3. These primers targeted mutations that were heterogeneously represented in the spatially separated tumour foci (Figure 4.1). TAm-Seq of both plasma and tumour tissue samples from case 7 and 8 were performed and libraries were pooled, cleaned and quantified, before being submitted with libraries from other projects and sequencing with 125bp paired end chemistry using a HiSeq 2500 (Illumina). Overall, 24.2 million reads aligned to the human genome for case 7 and case 8 samples combined. The median amplicon coverage was 5,700 and 11,000 for case 7 and 8 respectively with many primers failing to amplify the expected target loci, see Figure 4.13.

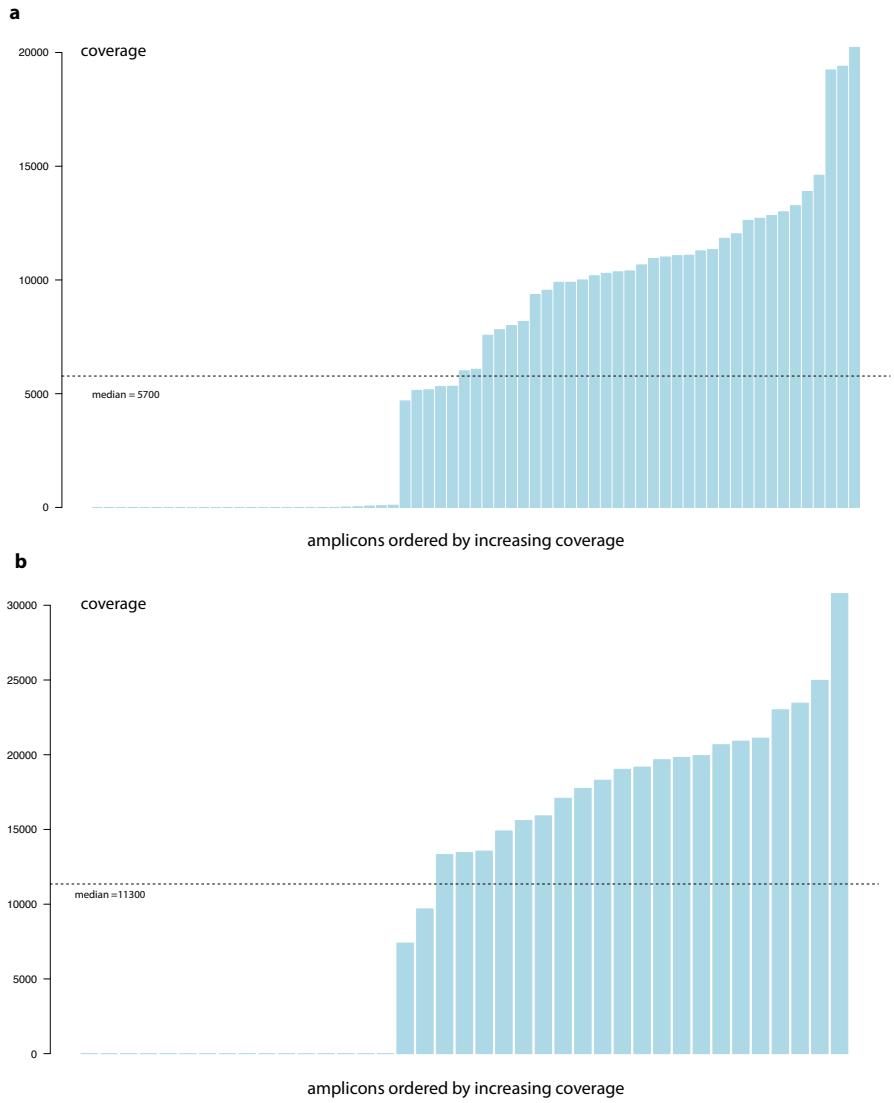



**Figure 4.13: Median coverage (y axis) per amplicon (x axis) with overall median coverage in dotted black for primer pairs targeting case 7 (a) and case 8(b), indicating that sequencing coverage was not even across the amplicons.**

Despite poor coverage in 30/70 and in 22/45 primer pairs for cases 7 and 8 respectively, available data was analysed for tumour AF concordance between TAm-Seq analysis, and WGS AF (supplied by the ICGC). Figure 4.14 demonstrates a good correlation between the two methods for the mutations tested ($r^2$ = 0.8-9.4). There were however, occasions when mutations were detected by one library preparation method only. Importantly, comparison of library preparation method with these results will be confounded by the failure of many primer pairs. When considering only primer pairs yielding sufficient depth, Figure 4.15a illustrates that 19 SNVs were detected by WGS that were not detected in TAm-Seq, and 9 SNVs were detected in TAm-Seq that were not present in WGS (for a specific sample), across all clones from case 7. These results suggest that while there is an overall good agreement between TAm-Seq and WGS library preparation methods, each method may have specific advantaged to detect individual mutations.

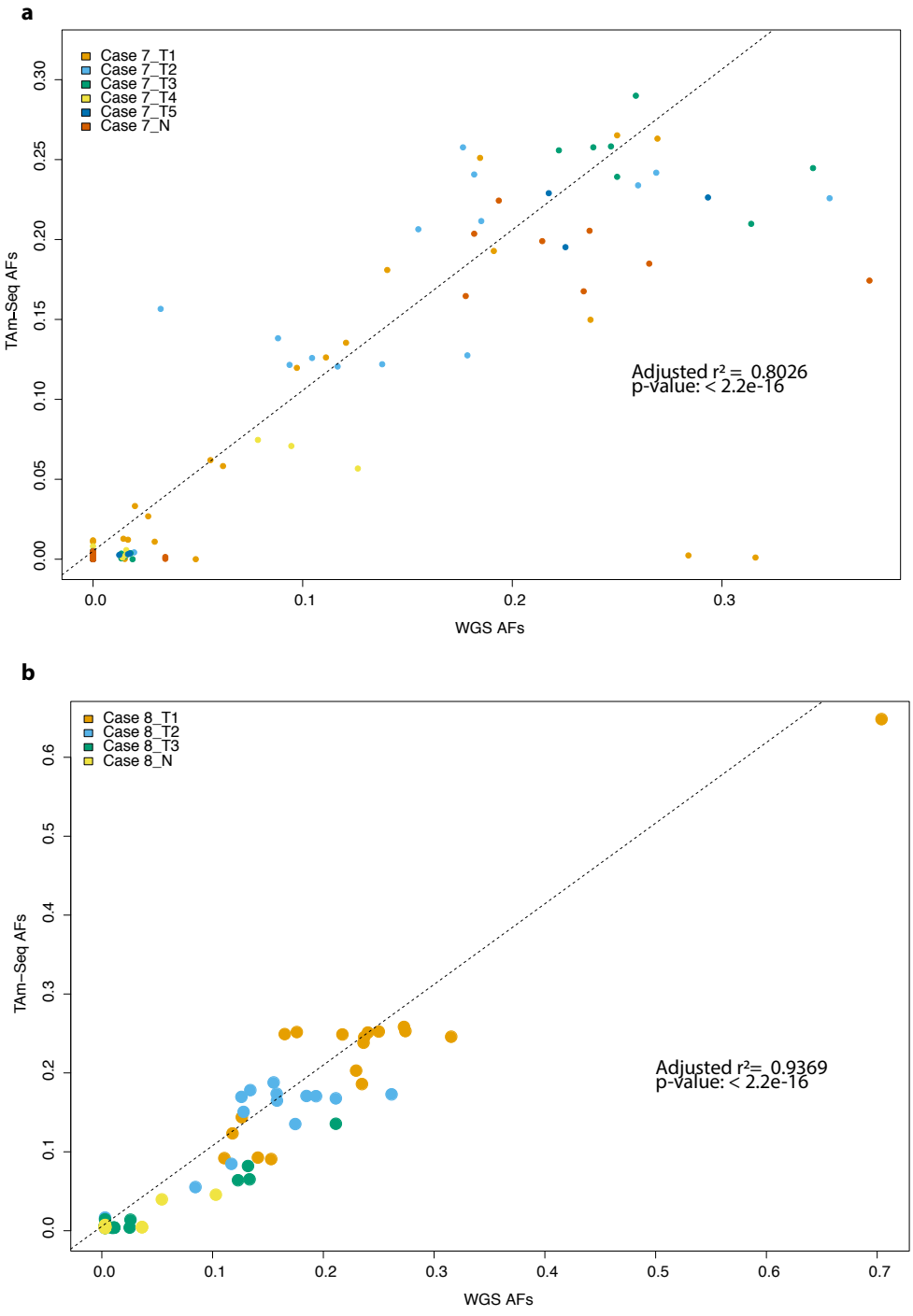



**Figure 4.14: Concordance of AFs between TAm-Seq and WGS determined AFs for Case 7 (a) and Case 8 (b).** Mutations identified in separate tumour foci are colour coded and linear modelling, using Wilkinson-Rogers model in R (R Core Team 2015), has been applied to assess correlation (dotted black line). Adjusted $r^2$ values demonstrate good agreement between the data types by the linear model for case 7 and 8 (0.8029 and 0.9369).

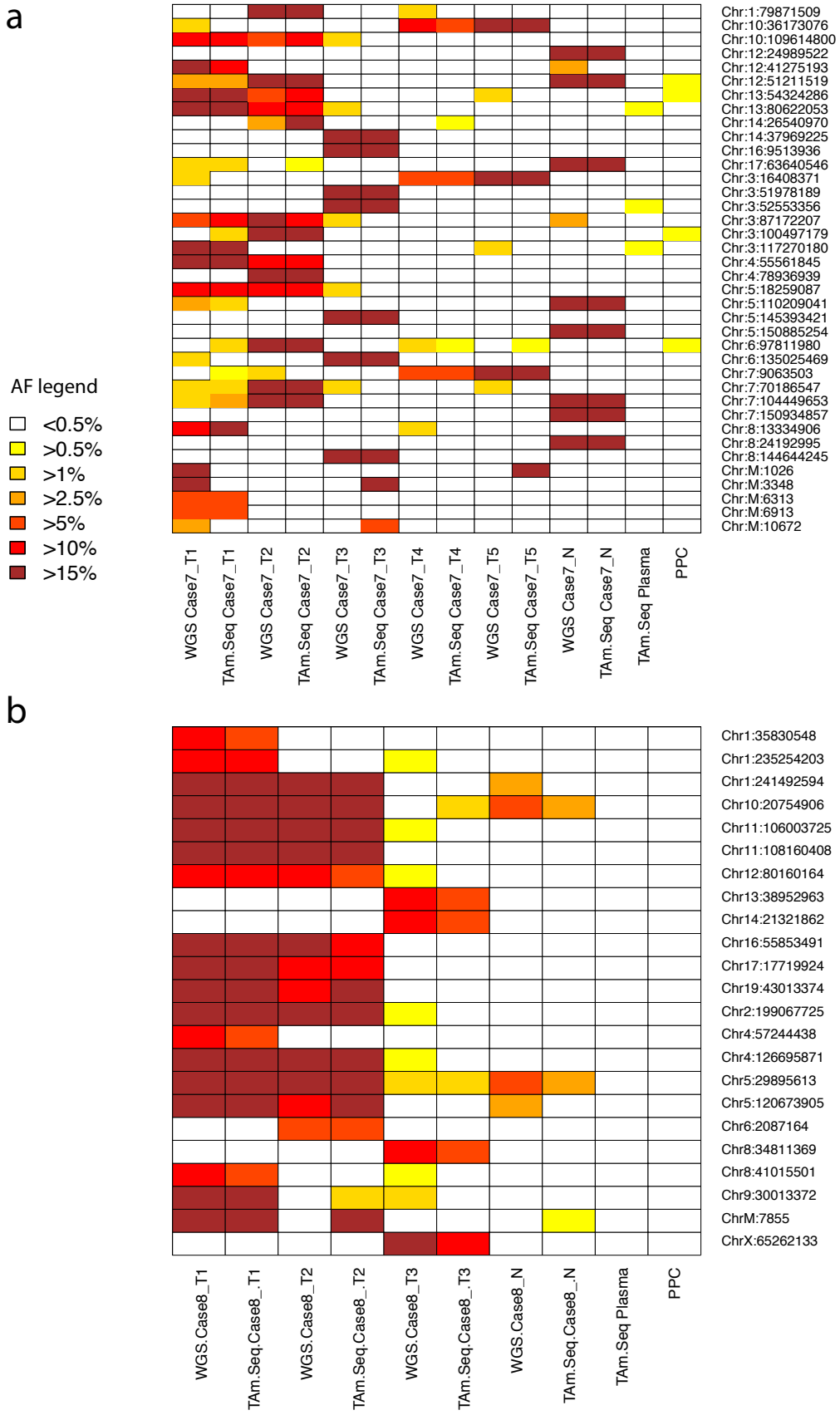



**Figure 4.15: Heatmap depicting AFs for each genomic location (y axis) in libraries prepared from spatially separate tumour foci and from plasma (x axis) in case 7 (a) and case 8(b)**

Table 4.14 shows the concentration of DNA extracted from the pre-prostatectomy plasma samples of case 7 and 8. Analysis of plasma AFs in detection mode (Forshew *et al.*, 2012) suggested ctDNA presence at low AFs but this was obfuscated by background noise (also present in some healthy control plasma samples, see Figure 4.15a). The inability to confidently detect ctDNA may be due to several reasons including; the loss of approximately half of all intended primer pairs or that levels of ctDNA are very low or absent in men with curable localised prostate cancer (both men were cured following surgery with no evidence of recurrence with >5 years of follow up).

| Sample ID | Extraction Sample Volume (ml) | elution Volume (ul) | Enrichment Ratio | CNA dPCR Loading Volume (ul) | Est. ROX Targets | Est. FAM Targets | Yield Est. | Amplifiable Copies/ml | Avearge Amplifiable copies /ml | Amplifiable copies / ul of eluate |
|---|---|---|---|---|---|---|---|---|---|---|
| Case 7 | 1 | 50 | 1 | 1.5 | 34 | 183 | 74.5% | 2340.6 | 2306 | 117 |
| Case 7 | 1 | 50 | 1 | 1.5 | 33 | 212 | 86.4% | 2271.7 | | |
| Case 8 | 1 | 50 | 1 | 1.5 | 60 | 203 | 82.7% | 4130.4 | 4130 | 209 |
| Case 8 | 1 | 50 | 1 | 1.5 | 60 | 186 | 75.8% | 4130.4 | | |
| PPC | 1 | 50 | 1 | 1.5 | 46 | 137 | 55.8% | 3166.7 | 3408 | 173 |
| PPC | 1 | 50 | 1 | 1.5 | 53 | 161 | 65.6% | 3648.6 | | |
| NEC_Plasma | 0.44 | 50 | | 1.5 | 0 | 251 | 102.2% | 0 | 0 | 0 |
| NEC_Plasma | 0.44 | 50 | | 1.5 | 0 | 240 | 97.8% | 0 | | |

**Table 4.14: Amplifiable GE copies in cfDNA extracted from plasma of men with heterogenous localised prostate cancer.** PPC – Healthy Plasma Control, NEC_Plasma – No Extraction Control.

### 4.5.7 Development of a Multiplexed Replicate Dilution Sequencing approach (MRD-Seq)

To further investigate the levels of ctDNA in localised prostate cancer a sensitive ctDNA detection process was developed to detect known mutations at low levels in cfDNA of bodily fluids. Despite improvements in the analytical sensitivity of methods, with the ability to detect mAFs as low as 0.01% reported (Vogelstein and Kinzler, 1999; Milbury *et al.*, 2014; Kinde *et al.*, 2011), the amount of available material remains a limiting factor for analysis of low levels of ctDNA. Methods that aim to detect a single mutation will have poor sensitivity to detect ctDNA at levels lower than one molecule per sample on average, even if they have excellent analytical accuracy (Bettegowda *et al.*).

One way to increase the number of mutant molecules sampled is to increase the amount of material collected, for example using large volumes of plasma by plasmaphoeresis (Marleau *et al.*, 2012). However these methods are under developed and would be challenging to implement in clinical practice. An alternative approach would be to increase the number of sampled mutant molecules, by analysis of multiple mutations in parallel (Newman *et al.*, 2014; Newman *et al.*, 2016). Methods that have been optimised for exquisite sensitivity (e.g. dPCR) have a limited capacity for multiplexing and generally

require samples to be distributed into a series of parallel reactions to achieve analysis of multiple targets (where rare molecules may not end up in the correct reaction) (Milbury *et al.*, 2014; Qlagen, 2016). On the other hand, methods that enable parallel analysis of multiple regions such as using next-generation sequencing have a limited sensitivity for each individual mutation, as illustrated by Forshew *et al.* (Forshew *et al.*). To overcome these limitations, we developed a method for genomics analysis of multiple regions whilst retaining sensitivity for detection of individual mutant molecules, using multiplex replicate dilution sequencing, or MRD-Seq.

MRD-Seq distributes a DNA sample into multiple replicate reactions, where each reaction contains a small number of initial template molecules. This number of template molecules per reaction ($N_w$), is determined based on the performance metrics of the sequencing process. The detection of a single mutant molecule in $N_w$ wild-type molecules will generate a sequencing signal that is clearly discernible above the background AF for non-reference reads in that sequencing process. For example, if the sequencing process has a background AF of 0.5%, $N_w$ can be chosen to be 40 molecules. Therefore, one mutant molecule in a pool of 40 molecules would create a sequencing signal with a mAF of ~2.5%, significantly higher than the background. Furthermore, each reaction can be multiplexed using primers that target mutations in multiple regions of the genome. The dilution process creates a "noise buffer" to limits false positive signals even if a large number of targets, $N_t$, is analysed and even if the multiplexed PCR amplification creates a high level of background sequencing noise. Each diluted replicate reaction can be separately barcoded, and a large number of replicates ($N_r$) sequenced. By sequencing a large number of multiple replicate dilution reactions, each including $N_w$ molecules, all available template molecules can be analysed ($N_m$=$N_r$x$N_w$). The process can analyse a large number of targets in each reaction, so that the total number of molecules sampled for the presence of mutations by the process, $N_s$=$N_r$x$N_w$x$N_t$, is much larger than the amount of template molecules for each mutation $N_m$, whilst reducing false positives. This approach would theoretically allow for sensitive detection of low levels of mutant DNA. With $N_w$=40, $N_r$=100, $N_t$=50, MRD-Seq could allow sensitivity down to a few mutant copies per 200,000 template molecules, even if the initial input included only 4,000 copies of the genome.

#### 4.5.7.1 MRD-Seq optimisation with high fidelity enzymes

**To improve background noise cause by polymerase errors during PCR reactions, Kapa HiFi HotStart (Kapa Biosystems KR0369), Accuzyme (Bioline BIO-21051), Phusion Hot Start II High-Fidelity (Thermofisher F549L) and Q5 Hot start (NEB M0493S) DNA polymerase enzymes were selected for**

**further testing based on error rates, hot-start and exonuclease activity from commercially available high fidelity enzymes, see**

Table 4.15.

| Enzyme | Kapa HiFi | Q5 Hot start | Phusion Hot Start II | Accu-zyme | Qiagen Hot Start | Pfu | Plati num | Roche | KOD |
|---|---|---|---|---|---|---|---|---|---|
| **Reported error rate vs Taq** | 100 | >100 | 50 | 47 | 10 | 8 | 6 | 4 | 4 |
| **Hot start** | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| **Exonuclease** | Y | Y | Y | Y | Y | N | N | N | N |

**Table 4.15: Comparison of high fidelity enzyme error rates, hot start and exonuclease activity.** Enzyme information was obtained from publications (McInerney *et al.*, 2014) and product technical datasheets literature.

Enzymes were subjected to the MRD-Seq workflow using healthy plasma template cfDNA (Sera labs) and the patient specific primer panels described above for case 7 and case 8 (Section 4.5.6.1). Each enzyme underwent PCR cycling using manufacturers recommended PCR reagents without additional DMSO or Magnesium Chloride. PCR cycling temperatures were agreed with respective technical support departments and using cycling timings matching TAm-Seq, including a long annealing time see Table 4.16. PCR products were analysed using DNA1000 chips on a Bioanalyser (Agilent).

| Enzyme | Temperature (ºC) | Time | No. of cycles |
|---|---|---|---|
| Kapa Hi-fi | 95 | 5 mins | 1 |
| | 98 | 20 secs | |
| | 60-72 | 4 mins | 20 |
| Q5 | 98 | 30 secs | 1 |
| | 98 | | |
| | 58-72 | 4 mins | 20 |
| Phusion | 98 | 30 secs | 1 |
| | 98 | 10 secs | |
| | 60-72 | 4 mins | 20 |
| Accuzyme | 97 | 3 mins | 1 |
| | 97 | 15 secs | |
| | 55-68 | 4 mins | 20 |

**Table 4.16: PCR cycling conditions used for assessing high-fidelity enzymes.**

Expected amplicon sizes were calculated based on the target sequence start and end locations with the addition of the primer pair tags that allow subsequent sample barcoding, see Figure 4.16.

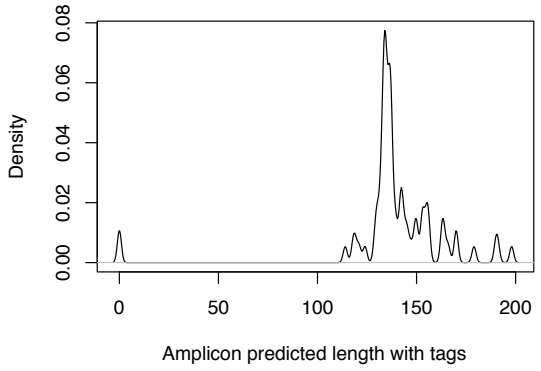
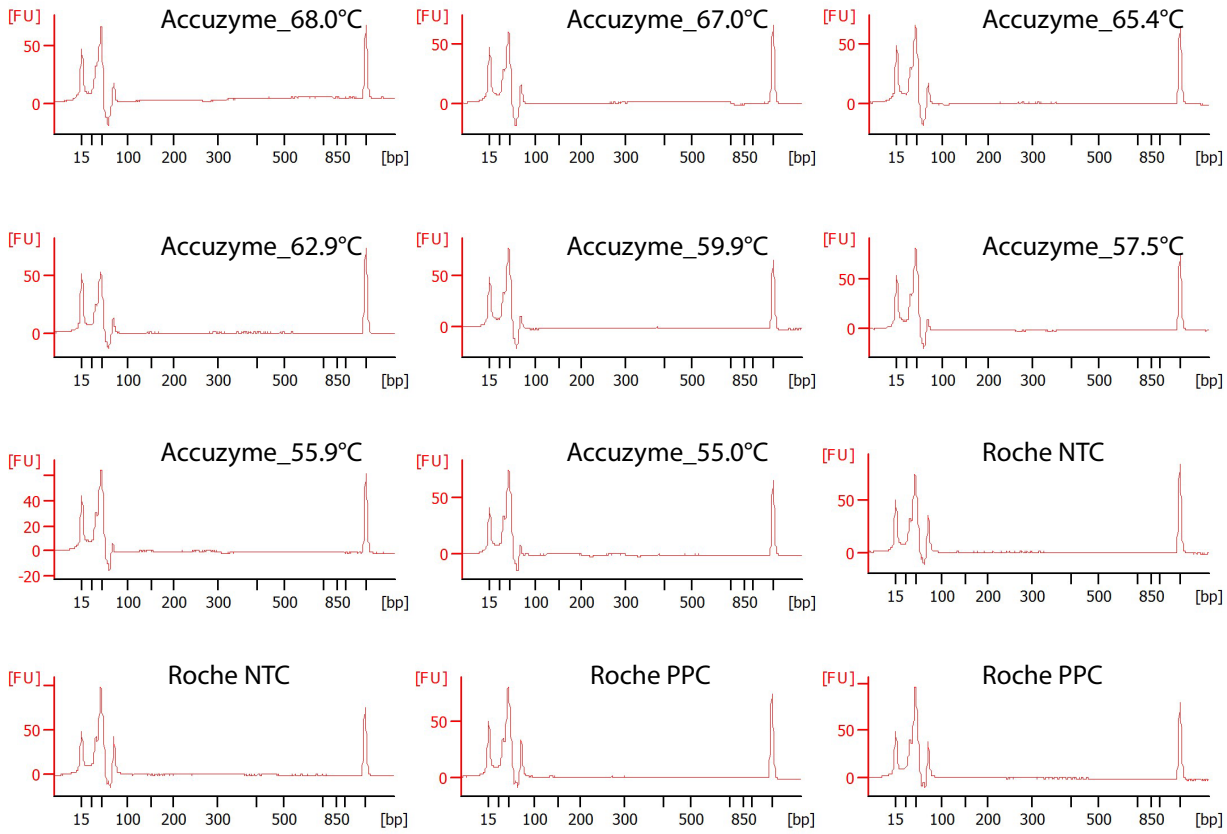



**Figure 4.16: Density plot of expected amplicon sizes with additional amplicon tags for the case 7 primer panel.**

Bioanalyzer traces of PCR products were produced to compare against the expected amplicon size range, see Figure 4.17. No trace followed the ideal size distribution, i.e. with a dominant peak at ~135 bases. Of note, following standard TAm-Seq pre-amplification, with samples taken from the experiments described in section 4.4.3, and using the Roche Polymerase, the Bioanalyzer trace had little or no product at the expected range (Figure 4.17). One possibility for the absence of products in the expected size range is that the

PCR reaction has failed, but this is unlikely as the Roche enzyme pre-amplification material yielded on-target sequencing data after the subsequent access array and barcoding steps. More likely is that during the multiplexed PCR reaction, an abundance of primer dimers are created along with some specific products. Indeed, the length of 2 primers with tags would be ~90bp (as observed in most traces in Figure 4.16). During the subsequent singleplex PCR reaction, the intended product is amplified more efficiently, yielding predominantly the specific product and little primer dimer. Therefore, I proceeded to select optimum pre-amplification temperatures for each high fidelity enzyme based on the representation of peaks at ~150bp and reduced peak at the expected primer dimer peak (as indicated in Figure 4.16). A final list of temperatures and enzyme combinations input into the access array are listed in Table 4.17.



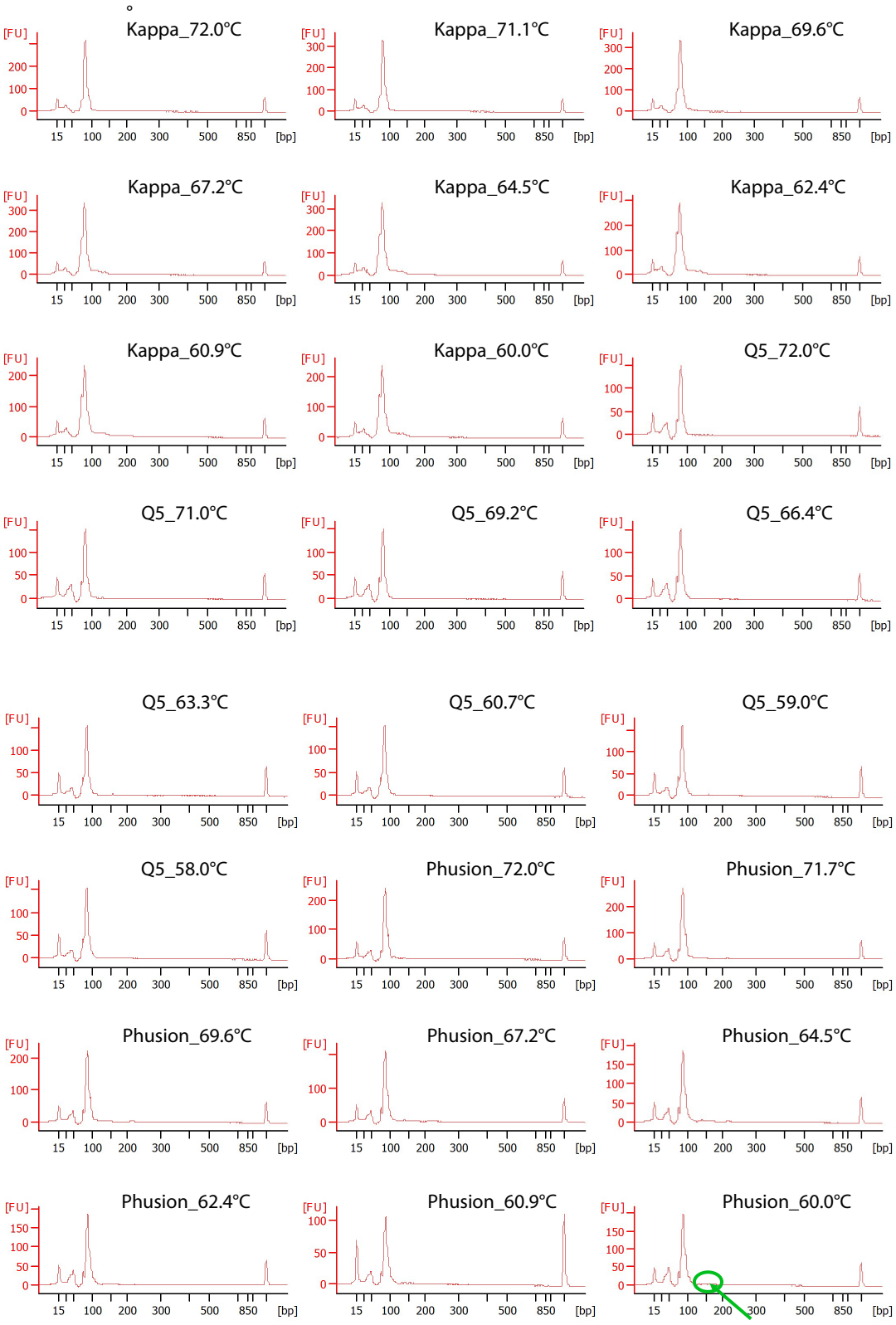



**Figure 4.17: DNA1000 Bioanalyser traces showing size distribution of PCR products created with high-fidelity enzymes using case 7 primer panel and template DNA extracted from healthy plasma.** Temperatures used for PCR cycling are recorded adjacent to the enzyme name. Expected size range is indicated in a trace (green arrow).

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| A | Kapa_63.3ºC | Kapa_63.3ºC | Kapa_63.3ºC | Kapa_63.3ºC | Kapa_63.3ºC | NTC |
| B | Kapa_63.3ºC | Kapa_63.3ºC | Kapa_63.3ºC | Kapa_63.3ºC | Kapa_63.3ºC | NTC |
| C | Q5_59ºC | Q5_59ºC | Q5_59ºC | NTC | Q5_59ºC | Q5_59ºC |
| D | Q5_59ºC | Q5_59ºC | Q5_59ºC | NTC | Q5_59ºC | Q5_59ºC |
| E | Phusion_64.5ºC | NTC | Phusion_64.5ºC | Phusion_64.5ºC | Phusion_64.5ºC | Phusion_64.5ºC |
| F | Phusion_64.5ºC | NTC | Phusion_64.5ºC | Phusion_64.5ºC | Phusion_64.5ºC | Phusion_64.5ºC |
| G | Accuzyme_67ºC | Accuzyme_67ºC | Accuzyme_67ºC | Accuzyme_67ºC | Accuzyme_67ºC | Accuzyme_67ºC |
| H | Accuzyme_67ºC | Accuzyme_67ºC | Accuzyme_67ºC | Accuzyme_67ºC | Accuzyme_67ºC | Accuzyme_67ºC |

**Table 4.17: List of temperatures and high-fidelity enzymes input into microfluidic multiple singleplex PCR.**

#### 4.5.7.1.1 Optimising PCR reactions for MRD-Seq

When starting with low numbers of molecules in MRD-Seq (e.g. 40 copies) each true variant would have a high AF (e.g. >2.5%) and so high depth sequencing may not be required. In addition the AccessArray singleplex step adds cost to the MRD-Seq process. To evaluate whether singleplex PCR was necessary for MRD-Seq, multiplexed PCR products listed in Table 4.17 were split into two groups. The first group were subjected to ExoSap cleaning and immediate barcoding, whilst samples in the second group were split to perform multiple singleplex PCR reactions using the Access Array chip (Fluidigm). Singleplex PCR thermocycling was carried out as per manufacturers recommendations. Samples from both groups were then barcoded as previously described (Forshew *et al.*, 2012). DNA libraries were purified, pooled and submitted for sequencing with 125bp paired end chemistry on the MiSeq platform (Illumina). Overall, 22 million reads were generated, however primer dimer accounted for the majority of reads, with aligned reads amounting to only 1.06 million, see Figure 4.18.

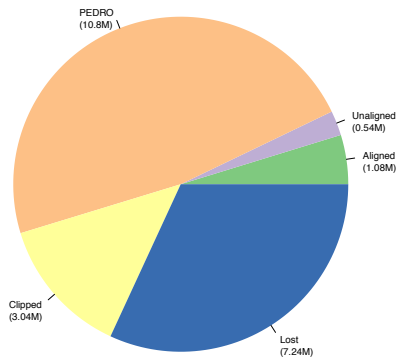



**Figure 4.18: MiSeq reads resulting from investigation of high-fidelity enzymes.**

Although raw read counts were broadly the same amongst the samples, reads aligning to the human reference genome (hg19) were not evenly distributed (Figure 4.19a). Samples undergoing secondary PCR with the Access Array produced aligned reads whilst the group of samples that omitted this step did not produce aligned reads (Figure 4.19 and Figure 4.20). Figure 4.19b shows that samples processed using the Q5 and Phusion enzymes accounted for the majority of aligned reads. Kolmogorov–Smirnov testing between Q5 and Phusion enzymes, revealed that Q5 samples had significantly better coverage ($p<2.2\ xe^{-16}$), see Figure 4.20.

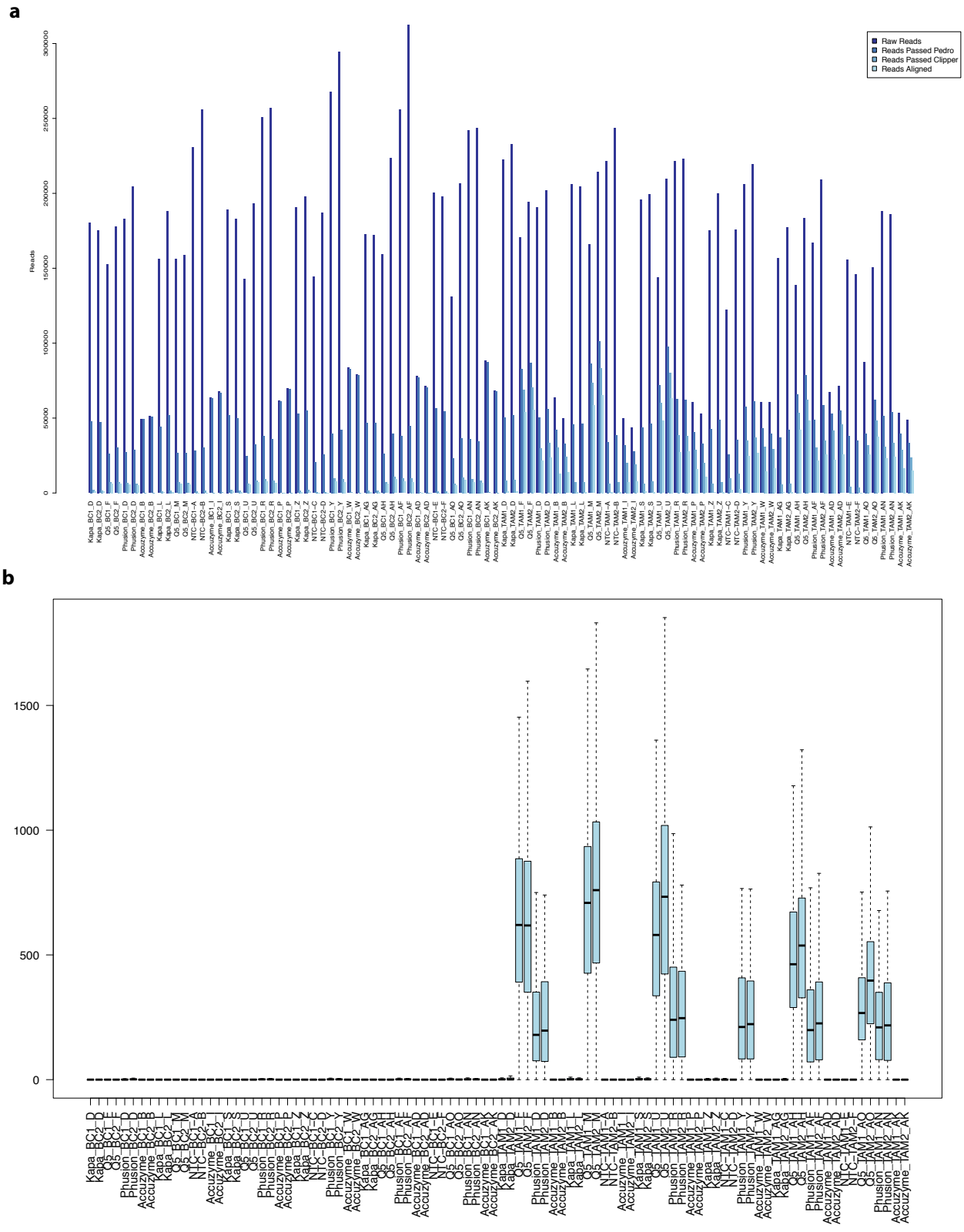



**Figure 4.19: Distribution of reads allocated to samples. a.** Proportion of raw reads, filtered reads and reads aligning for each sample. **b**. The median and interquartile ranges for coverage of aligned reads per sample.

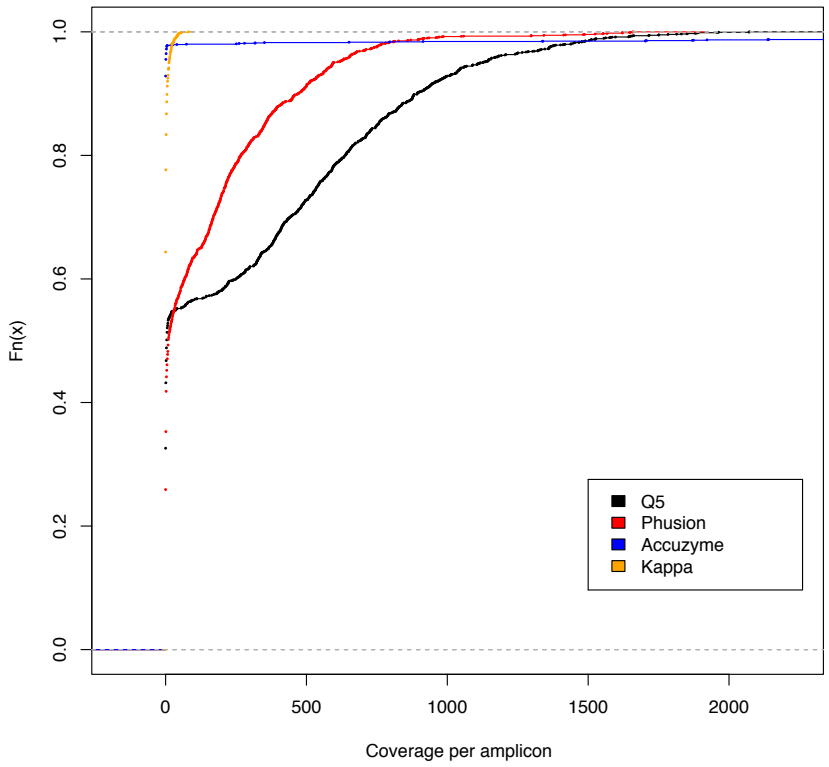



**Figure 4.20: ECDF plots of each group of samples undergoing PCR with a specified each enzyme type (coloured).** Empirical cumulative distribution function graphs depict the coverage (x axis) and the proportion of observations less than or equal to this value (y axis). Therefore ECDF plots can depict differences between the distributions data sets. One-sided Kolmogorov–Smirnov test demonstrated a significant difference between Q5 and Phusion enzyme coverage, confirming that Q5 samples had improved coverage ($p<2.2\ xe^{-16}$).

For the target genomic locations all non-reference read counts were compared to determine the background noise attributable to the Q5 and Phusion enzymes. Figure 4.21 shows the background noise profile for Q5 and Phusion enzymes in MRD-Seq replicates using 40 copies of healthy volunteer plasma DNA per reaction. This demonstrated a reduced background noise profile for Q5 over the Phusion enzyme, with fewer bases having an AF of 0.5% or more. Kolmogorov–Smirnov testing confirmed a statistically significant difference between distributions ($p=5.24\ xe^{-19}$). Therefore, the higher-fidelity Q5 enzyme utilising a second round of parallel singleplex PCR reactions was chosen for further optimisation.

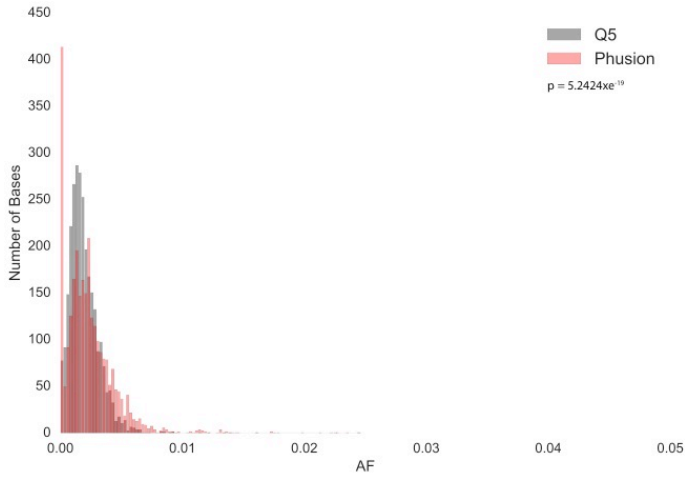



**Figure 4.21: Barchart showing the frequency of bases (y axis) having a non-reference AF (x axis) for MRD-Seq using healthy volunteer plasma.** Data generated using the Q5 and Phusion enzymes are shown.

#### 4.5.7.1.2 Optimisation of PCR cycling conditions for high-fidelity enzyme

To improve the yield of the desired products, PCR cycling conditions were optimised within the MRD-Seq workflow. The effect of magnesium concentration, enzyme concentration and annealing time on multiplexed PCR were explored. PCR annealing allows primers to attach to matched DNA template. During a multiplexed PCR, additional time given to this process can improve coverage but conversely can also increase the amount of non-specific product created e.g. primer dimer. Therefore, we tested 1-minute versus 4-minute pre-amplification annealing times. Katrin Heider (PhD student in Rosenfeld lab) had previously conducted experiments to improve the limit of detection for TAm-Seq using the Q5 polymerase enzyme. She showed that inputting 5 times the recommended Q5 enzyme concentration had significantly improved the yield of PCR product and therefore, we tested 1x and 5 x recommended enzyme concentration. Magnesium cation is required for DNA polymerase to function and its concentration can alter the efficiency of the enzyme. Higher concentrations will improve efficiency but lower the specificity of the enzyme (Markoulatos *et al.*, 2002; Bustin, 2010). Higher annealing and extension temperatures increase PCR product specificity but lower yield (Biassoni and Raso, 2014). Therefore we tested a annealing and extension temperature range from 58ºC to 70ºC.

The addition of DMSO has been shown to reduce nonspecific annealing and lowers the temperature of primer-template hybridization (Bustin, 2010). Therefore, it can be particularly useful in standardising the annealing temperature of multiplexed reactions especially when some primers are GC rich. However, the manufacturer recommends that

DMSO is not be required and provide a Q5 High GC Enhancer which can be used if PCR fails. Furthermore, primers were selected based on predicted annealing temperature and GC content. Therefore, the addition of DMSO was not tested.

DNA extracted from healthy volunteer plasma and case 7 patient-specific primer pairs were used. Samples were arranged in a 96 well plate before carrying out multiplexed PCR, using all case 7 primer pairs and the Q5 polymerase. Samples were subsequently transferred for parallel singleplex PCR using the Access Array (Fluidigm), barcoding, pooling and sequencing with 125bp paired end chemistry on the MiSeq platform (Illumina).

31.8 million reads were produced with only 0.83 million aligning to the reference human genome (hg19), see Figure 4.22. Again, the distribution of raw reads was broadly even amongst the samples, with a few samples accounting for the majority of aligned reads (resulting in a median depth of 25 across all conditions).

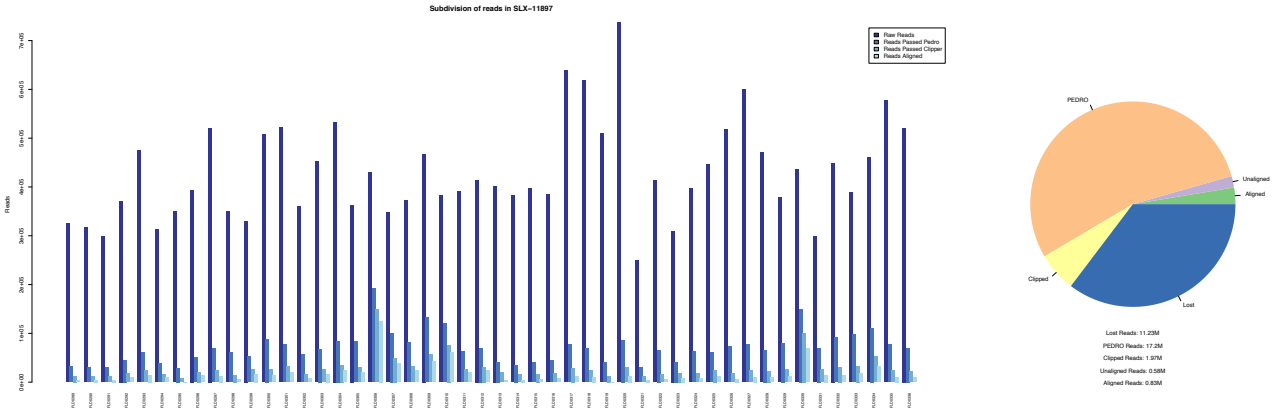



**Figure 4.22: Distribution of reads for the PCR condition optimisation experiment. a.** Boxplot of reads assigned to each sample, split by read class (raw, filtered, aligned). **b.** Pie chart of read allocations by read class across all conditions tested.

Figure 4.23 shows that a 2.5mM magnesium buffer concentration, recommended polymerase concentrations and an increased annealing time of 4 minutes, identical to the annealing time used in TAm-Seq (Forshew *et al.*, 2012), yields significantly higher target coverage (median depth of 77.5 for 2.5mM vs. 38 for 4.5mM magnesium buffer concentration, 14 for 1 minute vs. 148 for a 4 minute annealing time and 12 for 1x vs. 9 for 5x enzyme concentrations). Figure 4.24 demonstrates that a 2.5mM magnesium concentration and 1 minute annealing time had reduced background noise. There was no statistical difference between the error rate when varying the polymerase concentration.

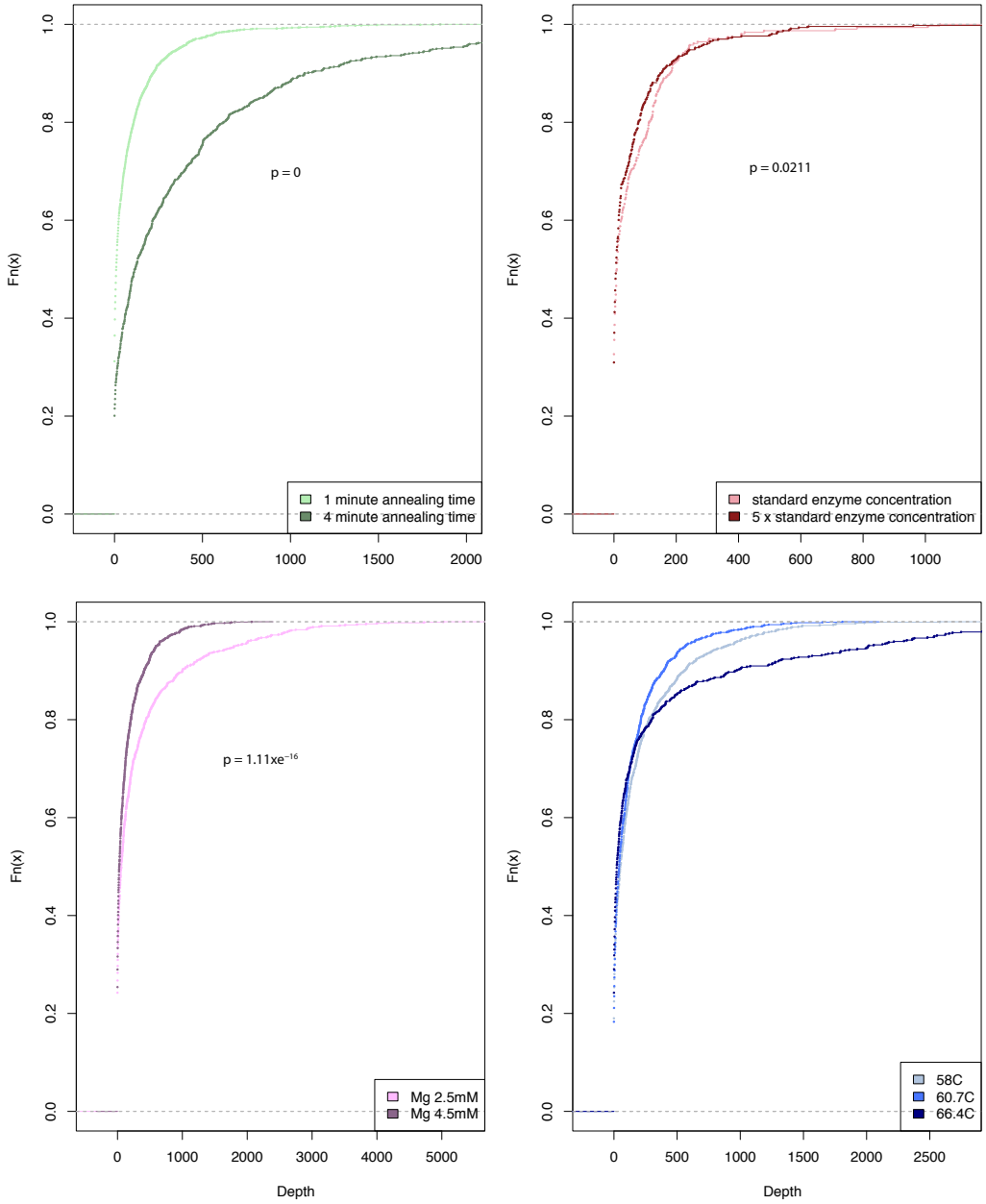



**Figure 4.23: ECDF plots of sequencing target depth for varying annealing time (a), polymerase concentration (b), magnesium concentration (c) and PCR annealing and extension temperatures (d).** Kolmagov-Smirnoff 2 sample testing was used to identify statistically significant differences in the distributions (inset p-values).

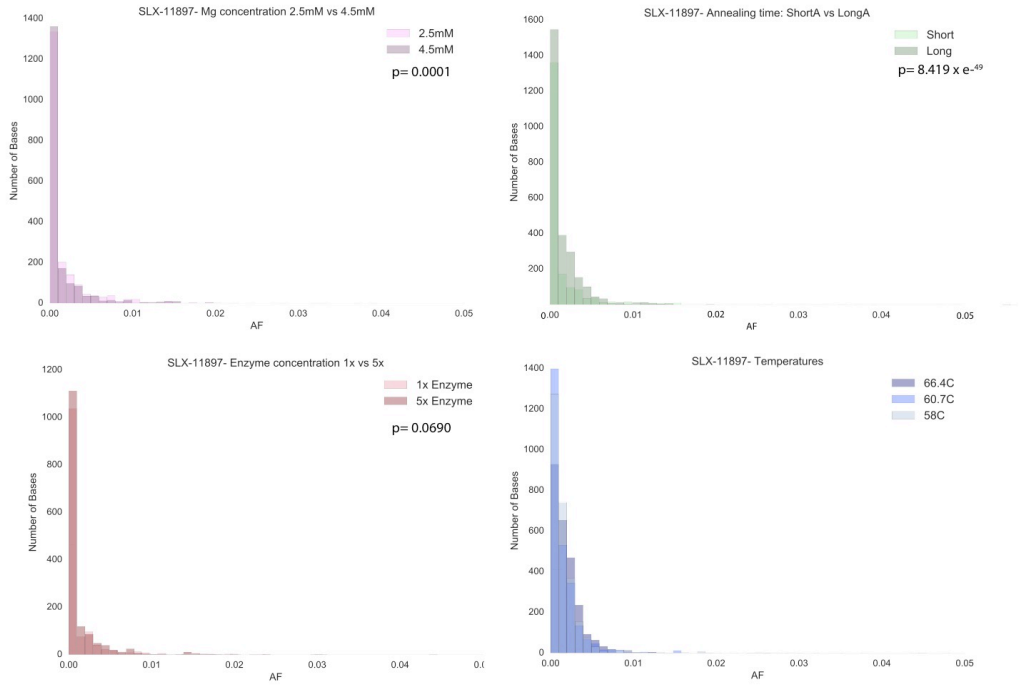



**Figure 4.24: Box plots comparing the background noise distributions from experiments comparing different Magnesium concentrations (a), annealing times (b), the polymerase enzyme concentrations (c) and PCR annealing and extension temperatures (d).** Kolmagov-Smirnoff testing was used to determine statistically significant distributions (noted as inset p-values).

Figure 4.23 and Figure 4.24 show that a magnesium concentration of 2.5mM increased depth and background noise level. The 4-minute annealing and extension time was maintained from the Tam-Seq protocol because results demonstrated that it improved DNA yield significantly (Figure 4.23) whilst having only a small, but statistically significant increase in background noise (Figure 4.24). Initial impressions, led to the use of 5x enzymes in the MRD-Seq workflow however, in-depth analysis reveals a small but significant improvement in DNA yield with standard polymerase concentrations (Figure 4.23).

Figure 4.23 shows that for optimal sequencing depth, a temperature of 60.7ºC should be used for Q5 PCR annealing and extension steps, with higher and lower deviations in annealing and extension temperature resulting in lower yields (Kruskal-Wallis p= 0.0033, p= 5.50xe$^{-7}$ and p= 3.930e$^{-06}$ for 58ºC versus 60.7ºC, 58ºC versus 66.4ºC and 60.7ºC versus 66.4ºC, after Bonferroni correction for multiple testing). However, Figure 4.24 demonstrates that higher annealing temperatures (66.4ºC) has a reduced background noise. To balance these factors, a compromised annealing and extension temperature of 63ºC was selected.

#### 4.5.7.1.3 Removal of primer dimer to improve proportion of aligned reads for MRD-Seq

The Rosenfeld group use a custom bioinformatics pipeline developed by Dr Francesco Marass and Dr James Morris to filter out primer-dimer reads (PEDRO). In brief, the pipeline filters primer-dimer sequences by removing reads where paired-end read 1 does not match the known primer pair of paired-end read 2. For data generated in section 4.5.6.1 with the case 7 primer panel using the standard TAm-Seq method, 9.9% and 35.7% of reads were filtered by PEDRO (1.89/19.1 million) and were aligned (6.82/19.1 million) respectively. During the optimisation of PCR conditions and reagents for MRD-Seq, 54.1% reads were filtered by PEDRO (17.2/31.8 million), whilst only 2.6% of reads were aligned (0.83/31.8 million).

Much of the 'off-target' reads were therefore attributable to short primer dimer artefacts. To investigate the effects of size selection to improve the proportion of on-target reads I performed DNA library size-selection using the Pippin Prep system (Sage). Pippin Prep uses an electrical gradient to separate DNA fragments along a gel matrix prior to eluting the selected size range in solution (Sage Science, 2017). I used the Pippin Prep system to remove PCR products in the primer-dimer size range from MRD-Seq libraries by selecting for DNA size fragments of between 105 – 200bp following pre-amplification or between 175 – 270bp following barcoding.

Libraries were generated using template DNA from healthy volunteer plasma and the optimised MRD-Seq workflow conditions (Section 4.5.7.1). Libraries were subjected either to size selection following pre-amplification (followed by Access Array parallel singleplex PCR, barcoding and pooling) or size selection after barcoding and pooling. Libraries were purified and submitted for sequencing using 125bp paired end chemistry with MiSeq (Illumina).

Figure 4.25a shows that 42.9% of reads were aligned following some form of size selection. Figure 4.25b shows that samples having size selection immediately following pre-amplification have a reduced proportion of reads removed by PEDRO and have a higher proportion of reads aligning to the human reference genome.

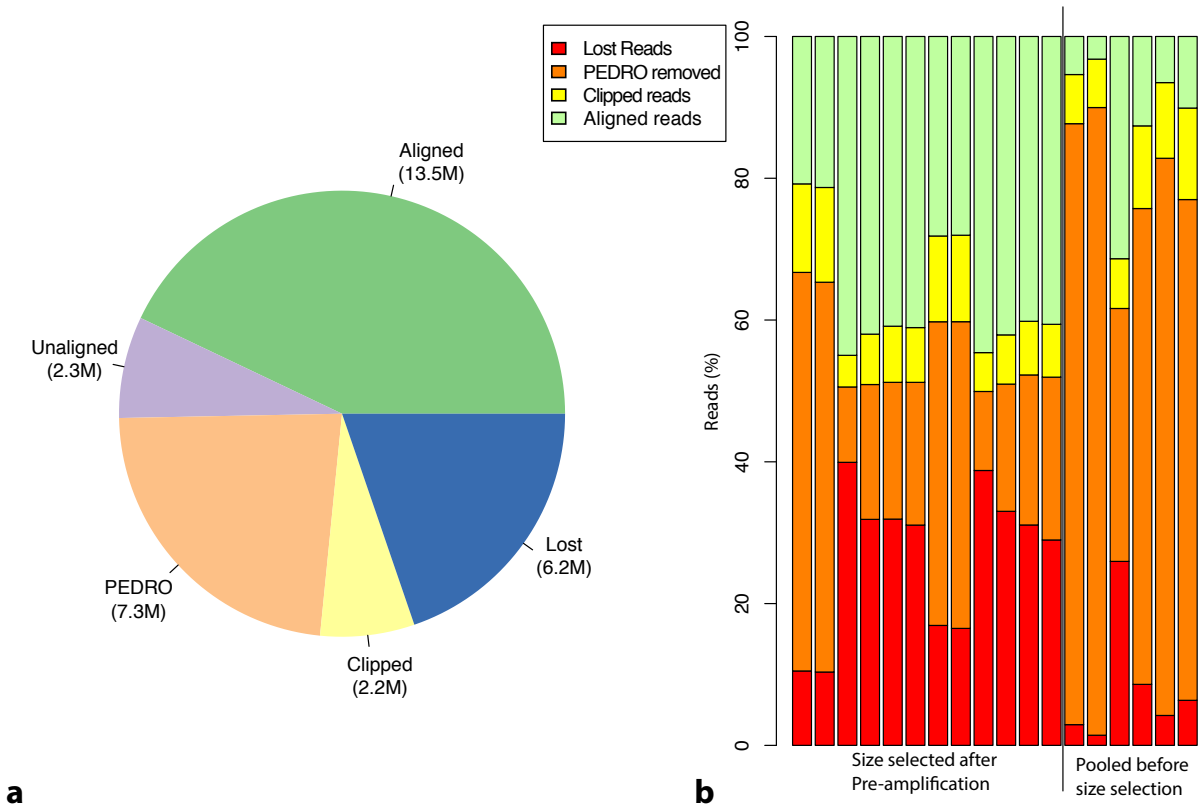



**Figure 4.25: Distribution of reads following library size selection. a.** Pie chart of read allocations by class**. b.** Barplot showing the percentage (y axis) of reads that were lost, removed by PEDRO, clipped and aligned for each sample, grouped by the point at which size selection was applied (x axis).

Library size selection after barcoding and pooling does increase aligned reads and would have streamlined the MRD-Seq workflow. However size selection immediately after the pre-amplification step increased the proportion of aligned reads dramatically and was therefore incorporated into the MRD-Seq workflow. Of interest, the results imply that the multiplexed pre-amplification step creates the majority of primer-dimers and hence size selection after this stage improves the proportion of reads aligning to the reference human genome.

### 4.5.7.2 Characterisation of MRD-Seq detection limit

To characterise the detection limit of MRD-Seq I performed a dilution series using tumour DNA diluted with wild type DNA (Promega, PR-G1471). Case 7 tumour region 2 DNA (Figure 4.2) was selected due to the presence of a high number of SNVs (14) at high AFs, 7 at ~20% AF and 7 at ~10% AF, see Figure 4.14. Bioanalyser traces showed that the DNA samples from tumour region 2 and Male genomic DNA (MCDNA) had similar size profiles, and dPCR was used to accurately quantify DNA stock concentrations, as described previously (see section 2.4.2). Table 4.18 demonstrates the dilution factors

assessed, dilution factors were calculated based on high AF (7 SNVs at ~20%). MRD-Seq was performed on samples according to the optimised protocol described above.

| Concentration ID | Expected AF for 14 mutations in region T2 | No of reactions ($N_r$) | No of GE copies / reaction ($N_w$) | No of Targets ($N_t$) | No of molecules sampled ($N_s$) |
|---|---|---|---|---|---|
| A | 0.156 | 4 | 40 | 14 | 2240 |
| B | 0.078 | 12 | 40 | 14 | 6720 |
| C | 0.0078 | 111 | 40 | 14 | 62160 |
| D | 0.0039 | 176 | 40 | 14 | 98560 |
| E | 0.0020 | 264 | 40 | 14 | 147840 |
| MC | 0 | 96 | 40 | 14 | 53760 |

**Table 4.18: Number of reactions ($N_r$) per dilution factor used to determine the detection limit of MRD-Seq.** Expected AFs were calculated by averaging the AF reported by WGS for the 14 mutations in case 7 region 2. MC- Male genomic DNA.

Overall, 663 reactions were performed. Samples were barcoded, cleaned and submitted for sequencing with 125bp paired-end chemistry over 2 lanes of HiSeq 2500 (Illumina). This resulted in 523.19 million reads, of which 37.6% aligned. The median depth for target loci was 90x. The expected AF for a true mutation in a single MRD-reaction was defined as 1/genomic input $N_{w,}$ using 40-copies as input resulted in a threshold of 2.5%. Replicates were deemed to collect one mutant molecule for every 2.5% AF recorded, e.g. 10% AF indicating the presence of 4 mutant molecules. Figure 4.26 depicts the AF at every mutant position tested across multiple replicates. There is a clear trend for AFs to reduce with increasing dilution factor. Table 4.19 demonstrates the overall AF estimate for each dilution. ECDF plots were generated (Figure 4.27) and Kolmogorov–Smirnov testing used to identify statistically significant differences between each dilution series and control DNA ($p<0.0001$ for all pairs).

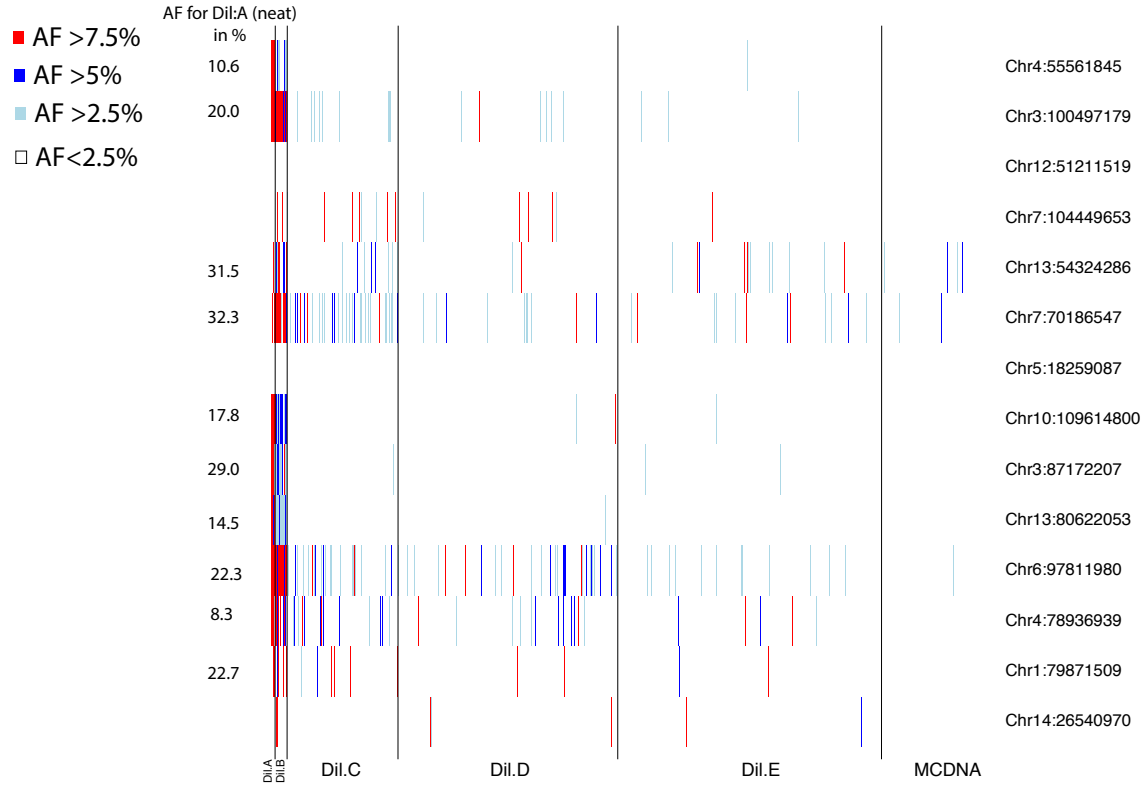



**Figure 4.26: Heatmap demonstrating the number of molecules detected during characterisation of MRD-Seq sensitivity.** Replicates are grouped by dilution factor (x axis). Mutant position locations are recorded on y axis. AFs for Dilution A shown to the left of the y axis.

| Concentration ID | Expected AF for 14 mutations in region T2 | No of mutant detected ($N_m$) | No of molecules sampled ($N_s$) | Overall AF estimate using MRD-Seq |
|---|---|---|---|---|
| **A** | 0.156 | 203 | 2240 | 0.0906 |
| **B** | 0.078 | 426 | 6720 | 0.0634 |
| **C** | 0.0078 | 2280 | 62160 | 0.0367 |
| **D** | 0.0039 | 153 | 98560 | 0.0016 |
| **E** | 0.0020 | 135 | 147840 | 0.0009 |
| **MC** | 0 | 13 | 53760 | 0.0002 |

**Table 4.19: Characterisation of the MRD-Seq detection limit.** Expected AFs were calculated as described in **Table 4.18**. Overall AF estimates using MRD-Seq for each dilution were calculated by dividing the number of mutant molecules detected by the number of molecules assessed ($N_m/N_s$).

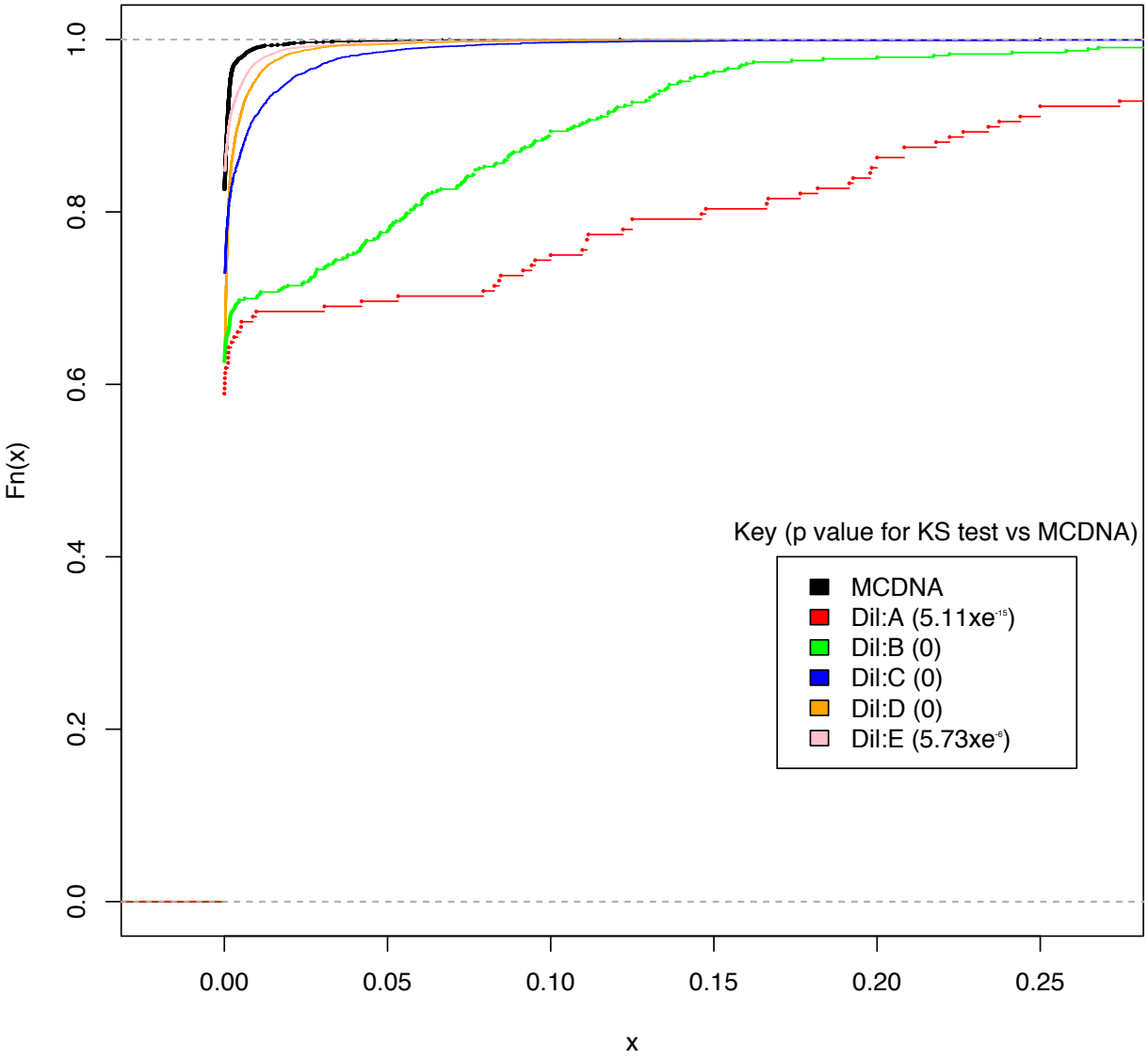



**Figure 4.27: ECDF plot demonstrating cumulative AF distribution for each dilution to determine the limit of detection for MRD-seq.** AFs are depicted on the x axis, with the proportion of samples with AFs less than that value are depicted on the y axis. Samples/dilution factors are grouped by colour code (see legend) and one-sided Kolmogorov–Smirnov test p-values are inset adjacent to the sample IDs (see legend). Statistically significant differences were demonstrated for each dilution factor and the healthy control DNA (MCDNA) distribution.

#### 4.5.7.3 Utilising MRD-Seq to detect ctDNA AFs in heterogeneous localised prostate cancer

To demonstrate the effectiveness of this process, I applied MRD-Seq to quantify ctDNA levels in the plasma of two men with multifocal localised prostate cancer (man 7 and 8), where 4-6 regions of prostatectomy samples underwent whole genome re-sequencing, as described previously (Cooper *et al.*, 2015), see section 4.4.3. One of these regions included an area of adjacent, histo-pathologically normal area for each man. From WGS data I selected 88 and 78 mutations for man 7 and 8 respectively to represent the intratumour heterogeneity of each man's tumour. Of these 69 and 45 assays were designed for man 7 and man 8 (number of targets, $N_t$) on the basis of high mAF in the tumour. Pre-operative plasma from man 7 and 8 had 2300 and 4130 amplifiable GE copies/ml of DNA respectively (total cfDNA). Replicate dilutions were performed from approximately 1000 GE copies such that the number of template molecules per reaction ($N_w$) *is* 38 and 39 for man 7 and man 8, respectively.

Detection thresholds ($D_t$) were calculated by defining a ratio of one mutant copy over the total GE input for each reaction (see Table 4.20) to give an expected lower limit of ~2.5% AF for both men. Healthy volunteer plasma (HVP, Seralabs) was used to define the background error rate for the target mutations. DNA was extracted from HVP using QIAsymphony (QIagen) and quantified with dPCR. HVP cfDNA replicates were made such that 40 copies were aliquotted into each reaction ($N_w$=40) in the MRD-Seq process and were analysed in parallel to the MRD-Seq plasma samples. Illumina sequencing was performed on pooled libraries (MiSeq for controls and HiSeq for plasma).

Figure 4.28 shows that median coverage for each amplicon was variable amongst sample types and in particular, demonstrates poor coverage for the majority of amplicons from the pre-operative plasma DNA.

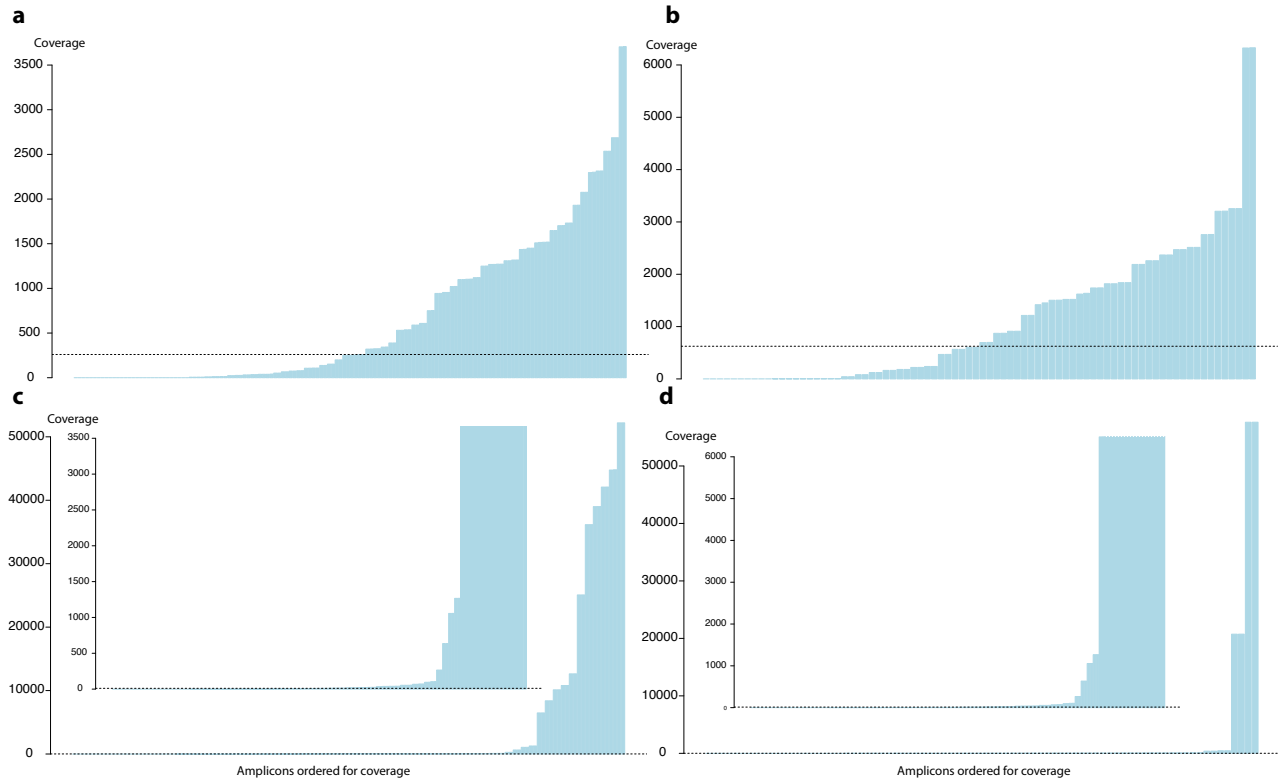



**Figure 4.28: Median amplicon coverage for MRD-Seq from healthy volunteer plasma (HVP) samples with man 7 assays (a), pre-operative plasma from man 7(c), HVP samples and man 8 assays (b) and pre-operative plasma from man 8(d).** Bars represent the median coverage (y axis) for each amplicon in the panel (x-axis). The dotted line represents the median coverage across all amplicons for a given sample (a= 260, b= 626, c= 10, d= 11). Inset figures within **Figure 4.28b and d** represent the coverage with a limited y-axis.

Figure 4.29 shows the distribution of plasma AFs for each man compared with HVP and demonstrates that there is no statistically significant difference between the two distributions for either man. However, the analysis may be confounded by the reportedly lower error rate of HiSeq (used for the patient samples) compared to MiSeq (used for HVP) sequencing (Quail *et al.*, 2012). Indeed, while KS testing demonstrated no significant difference between the AF distributions, the AFs for control samples, sequenced on a MiSeq platform, appear to be higher than plasma AFs. A more specific assessment of signal in this experiment may therefore be to count events above the expected detection threshold ($D_t$).

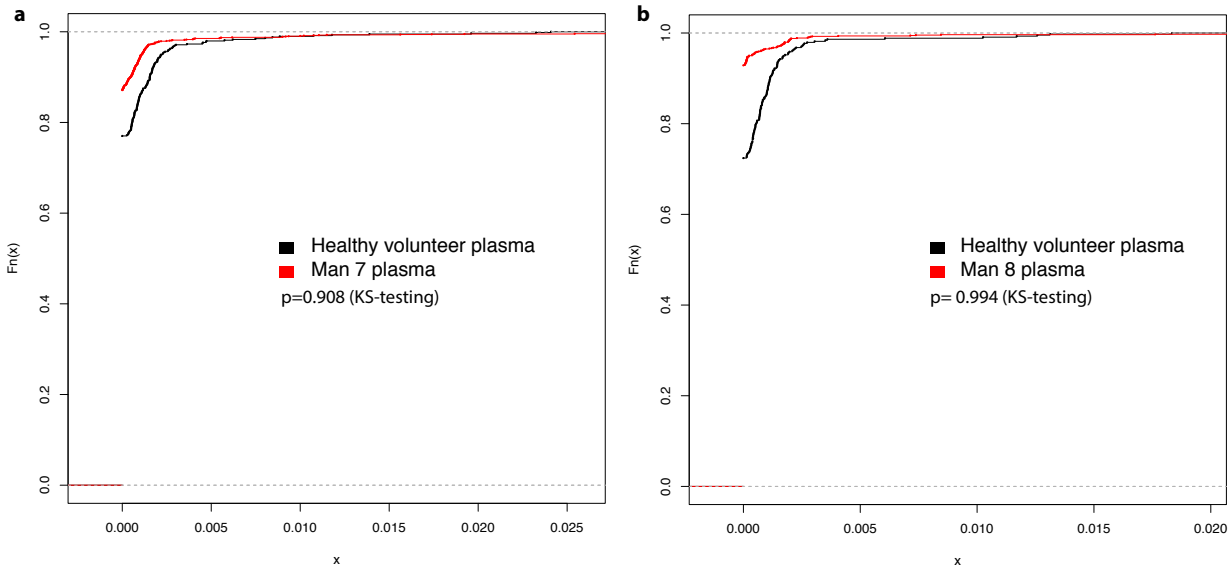



**Figure 4.29: ECDF plot showing the distribution of AFs in the plasma (red) and HVP (black) replicates of man 7 (a) and man 8 (b).** HVP – healthy volunteer plasma. One-sided Kolmagov-Smirnoff testing did not demonstrate statistically significant difference in the distributions of neither man (noted as inset p-values).

| Pt ID | dPCR GE Copies /µL of Eluate | GE Copies / reaction | Replicates Used | Nucleotides Sampled ($N_s$) | Mutant Molecules Identified | Estimated AF |
|---|---|---|---|---|---|---|
| 7 | 46.1 | 40 | 24 | 818 | 8 | 0.009 |
| HVP (for man 7 assays) | 89.7 | 40 | 14 | 614 | 0 | 0 |
| 8 | 82.6 | 40 | 24 | 484 | 2 | 0.004 |
| HVP (for man 8 assays) | 89.7 | 40 | 14 | 395 | 0 | 0 |

**Table 4.20: Input amounts GE copies, assays targeted, number of replicates, mutant molecules and sensitivity estimates for each patient.** HVP – Healthy Volunteer Plasma.

Sequencing reads were filtered for depth resulting in 818 and 484 molecules sampled ($N_s$) for pre-operative plasma and 614 and 195 molecules sampled for HVP samples for man 7 and 8 respectively, Table 4.20. Heatmaps were created to assess the AF for every depth-filtered amplicon, highlighting the AFs that crossed the MRD-Seq detection threshold (Figure 4.30). Of note, there were no healthy control plasma molecules with a mutant AF above $D_t$ (0 false positives). However, 8 and 2 mutant molecules above $D_t$ plasma were

demonstrated in man 7, and 8 respectively (Table 4.21) and were present in more than one read and on both the positive and negative strand.

| Case | Mutation | Gene | Effect | Clonal origin | Mutant Reads (AF%) |
|---|---|---|---|---|---|
| 7 | chr3:100497179 G>A | ABI3BP | p.R717* | T2 | (6.25) |
| 7 | chr7:104449653 C>A | LHFPL3 | Intronic | N+T1+T2 | (14.29) |
| 7 | chrM:1968 G>A | | | T1 | (5.26) |
| 7 | chr8:24192995 G>A | ADAM28 | p.D470N | N | (5.00) |
| 7 | chr11:41199322 A>G | | Intergenic | T2 | (10.53) |
| 7 | chrM:5175 C>T | | Intergenic | T1 | (2.54) |
| 7 | chr6:74612812 G>T | | Intergenic | T3+T4+T5 | (3.70) |
| 7 | chr3:87172207 C>A | | Intergenic | T1+T2 | (3.13) |
| 8 | chr16:55853491 G>A | CES1 | p.R288* | T1+T2 | (5.88) |
| 8 | chr14:21321862 T>C | | Intergenic | T3 | (3.85) |

**Table 4.21: MRD-Seq results showing mutation position, gene, effect, clonal origin and AFs for mutant molecules above $D_t$.**

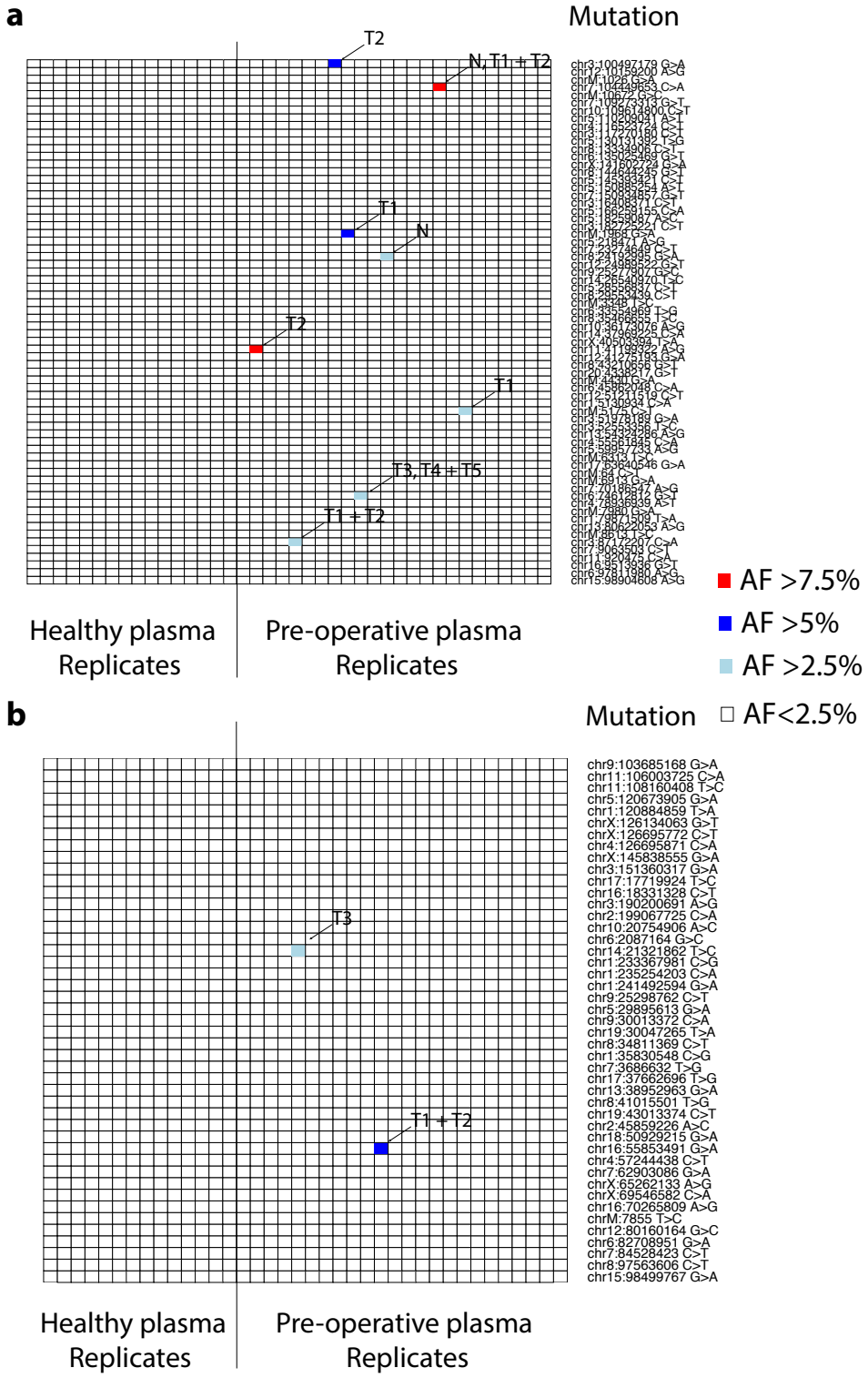



**Figure 4.30: Heatmaps showing the distribution of mutations detected in MRD-Seq experiments.** The x-axes denotes the replicate reactions, whilst the y-axes shows the genomic targets (mutations). AFs for each sample:location are colour coded according to patient specific detection thresholds illustrated in the legend. Healthy volunteer re-sequencing results are display on the left of the heatmap for comparison. **Figure 4.30a** and **Figure 4.30b** show the results for man 7 and 8. The MRD-Seq replicates detected signal from 8 molecules with AF>2.5% for man 7 (**a**) and 2 molecules with AF>2.5% from man 8 (**b**).

To assess the distribution of signals for each mutation independently, the AFs for all replicates were plotted on barplots. Figure 4.31 shows examples of AFs at specific genomic locations. These plots demonstrate that at many positions, MRD-Seq replicates did not detect a mutant molecule (i.e. all AFs were below $D_t$; Figure 4.31a and b). However, for a subset of target positions the AF of an MRD-Seq replicate was clearly above $D_t$. Figure 4.31c and d show examples of replicates with an AF significantly above the AF in other replicates / healthy controls at the same location. Although for each mutation position only one MRD-Seq replicate crosses the $D_t$ limit, the detection of mutations in patient samples and their absence on HVP controls suggest the presence of mutant ctDNA molecules in these early stage prostate cancer plasma samples.

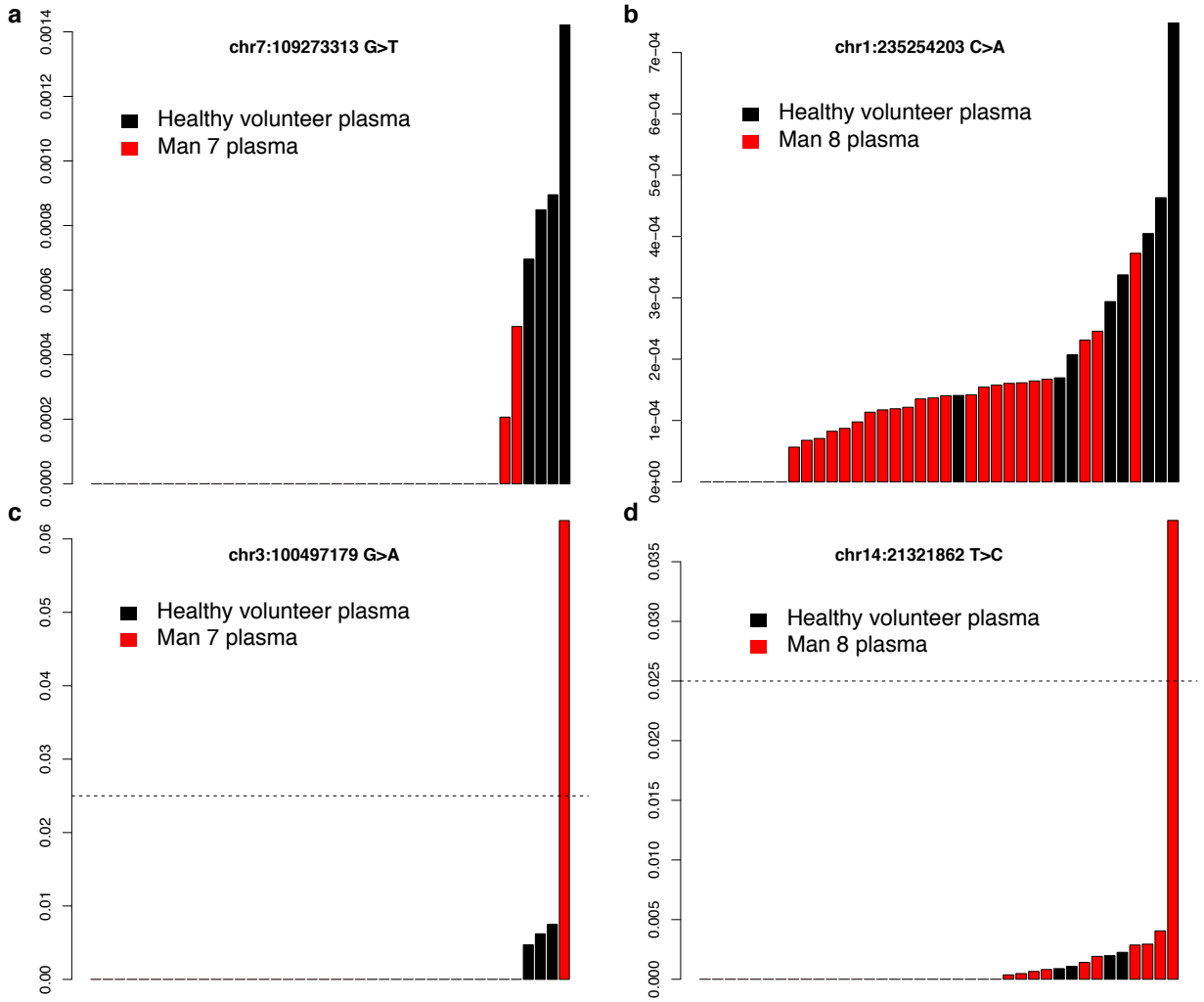



**Figure 4.31: Waterfall plot showing the ordered Allele Fractions (y axis) for patient (red) and healthy control (black) plasma replicates (x-axis) at four example genomic locations. a** and **c** shows AFs for man 7 and **b** and **d** shows data for man 8. $D_t$ is represented by a dotted line.

Despite the uneven sequence coverage in these pilot experiments, MRD-Seq analysis suggests the presence of 8 mutant molecules out of the 818 molecules tested in man 7 and 2 mutant molecules out of 484 for man 8, demonstrating AFs of approximately 1 in 150 molecules or 0.7% in localised prostate cancer (see Table 4.20 and Figure 4.30). Of potential clinical importance, the mutations detected in plasma ctDNA were enriched for specific regions in the prostate of these men. Specifically, plasma from man 7 contained, 2 mutant molecules emanating from region T1, 2 from region T2 and 1 from region N (but none from regions T3, T4 or T5). Three further mutant molecules represent shared mutations from regions T1/T2, regions N/T1/T2, and regions T3/T4/T5 respectively. Therefore, 7/8 mutations detected in ctDNA are derived from one half of the predicted genomic phylogeny (i.e. regions T1/T2/N and not regions T3/T4/T5), suggesting a bias for ctDNA release from tumour clones in these regions. Plasma from man 8 contained clones emanating from region T3 and a shared mutation from regions T1 and T2. (Figure 4.2 and Figure 4.30). This proof-of-principle study suggests that ctDNA analysis of plasma may demonstrate the complexity of heterogeneous localised prostate cancer but requires further investigation. If corroborated in larger cohorts, ctDNA analysis could be used to monitor tumour clones for the appearance of aggressive mutations (e.g. *TP53*) and alert clinicians to the presence of significant disease at an early stage.

## 4.6 Original Contributions to Knowledge

1. We confirmed that *TP53* SNVs are more prevalent in metastatic prostate cancer. Of potential clinical significance, we found that these SNVs could be detected in historical samples taken at the time of prostatectomy.
2. We developed MRD-Seq to demonstrate private mutations in ctDNA from multifocal prostate cancer in 2 cases.

This data suggests that ctDNA can detect private mutations from multiple clone and therefore may be able to detect the emergence of potentially aggressive mutations e.g. *TP53*.

# CHAPTER 5: DISCUSSION

## 5.1 Urinary DNA extraction

Since the conception of the project, other groups have detected mutDNA in urine. In 2016, Togneri *et al.* compared the analytical sensitivity of mutant detection using Oncoscan (Affymetrix) and DNA extracted from urine cell pellets and urine supernatant at diagnosis from patients with bladder cancer (Togneri *et al.*, 2016). They utilised the Urine DNA Isolation Slurry kit (Norgen) to extract DNA from an average of 18ml of urine. Unfortunately, urinary cfDNA yield or quantification data was not provided. However, that there was adequate DNA material for the 12ng input into the Oncoscan infers that DNA yields were adequate, lending support to results that column based extraction methods are suitable for urinary mutDNA analysis. Also in 2016, Birkenkamp-Demtröder *et al.* reported the use of the QIAsymphony Circulating NA kit (Qiagen, Hilden, Germany), to extract cfDNA from approximatley 3ml of 101 urine supernatants, ranging from 0.8ml – 4.5ml (Birkenkamp-Demtroder *et al.*, 2016). Reported purification efficiencies ranged from 1.3% - 100% with a pooled mean of approximatley 51.8% (calculated accross reported means for each group of patients) and copies/ml ranged from 0 – 654,872, with a pooled mean of 21,579 copies/ml. Birkenkamp-Demtröder *et al.'s* reported mean is almost identical to the mean we report (22,810 GE copies/ml).

In our group of 7 volunteers, of the four DNA extraction methods tested, SnoMag (which has since ceased production) was the only kit to have statistically significantly reduced cfDNA yields. However the small number of volunteers and that one commercial DNA extraction kit producer has potentially changed their extraction kits will impact statistical analysis. Furthermore, that urinary cfDNA can be collected in large volumes could negate the need for excellent yield of extraction and make other features of extraction, e.g. size of cfDNA extracted or reliability more important. Comparisons of urinary cfDNA size between the four extraction methods used could have been assessed by sWGS. However, on balance, it was felt that the Qiagen circulating nucleic acid kit would be the most suitable for further experimentation due to the Rosenfeld laboratory's significant experience with using Qiagen circulating nucleic acid kits for plasma DNA extraction, the higher mean yield of DNA when testing this kit with urine and the lack of statistical evidence supporting the use of another kit.

Since the testing of urinary cfDNA extraction, several other companies have released commercially available kits with specific urinary cfDNA extraction protocols. For example, NEXTprep-Mag™ Urine cfDNA Isolation Kit (Bio Scientific) have released a magnetic bead based cfDNA extraction method capable of DNA extraction from 1 – 50ml of urine. MagMAX™ Cell-Free DNA Isolation Kit (ThermoFisher Scientific) have also released a magnetic bead based method capable of extracting cfDNA from 1 – 10 ml or urine supernatant. Sigma-Aldrich offer the column based GenElute™ Urine Cell-Free DNA Purification Mini Kit, isolating cfDNA from 250μL-2mL of urine supernatant. Furthermore, several groups report using plasma DNA extraction kits to extract urinary DNA, for example, Tsui et al. report using Wizard PLUS kit (Promega) to investigate foetal and maternal urinary DNA. These and other urinary cfDNA extraction methods have not been investigated as part of this project, though could form part of future work if deemed appropriate.

### 5.1.1 Urinary processing protocols

Our data suggest that timely urine processing, i.e. within 1 hour, with centrifugation and the addition of EDTA urine improves cfDNA yield. When considering centrifugation of urine. for non-urological cancers, this would seem vital to reduce contamination of wild-type DNA from shed urothelial cells however, no comparisons have been published. Though no comparison was reported, Togneri *et al.* and Birkenkamp-Demtröder *et al.* describe the use of urinary centrifugation with 2000rpm and 3000g respectively for 10 minutes for urine processing in their analysis (Togneri *et al.;* Birkenkamp-Demtroder *et al.*). However, for urological cancers, particularly BC, one might consider un-centrifuged or whole urine analysis. Indeed, whole urine may allow investigation of private UCP and USN mutants in one sample type with reduced processing requirements (potentially allowing sample processing in peripheral hospitals) and would be worthy of further analysis. The addition of EDTA was not strongly supported by the results of our analysis. Furthermore, neither Togneri nor Birkenkamp-Demtröder *et al.* report using EDTA in urine for mutDNA analysis. However, for urinary cfDNA analysis several reports utilise EDTA (Tsui *et al.*, 2012) and EDTA has been shown to improve DNA stabilisation in plasma over other stabilisers (Bronkhorst *et al.*, 2015) or no stabiliser (Barra *et al.*, 2015).

Many studies use a sample processing time of 1-4hours for the analysis of ctDNA in plasma. Our data corroborates that there is a significant difference in DNA yields when samples are left unprocessed for longer periods of time. These findings suggest that

obtaining samples from peripheral sites would be problematic not least due to the requirement for high-speed centrifuge. However, Cell-Free DNA BCTs (Blood collection tubes, Streck) have been shown to improve DNA yield in plasma stored at room temperature for several days (Norton *et al.*, 2013) and outperform the addition of EDTA over this time period for the detection of ctDNA (Parpart-Li *et al.*, 2017; Sherwood *et al.*, 2016). Streck also promote the Cell-Free DNA Urine Preserve tube for the preservation of 25-100ml of urine for up to 7 days at room temperature and have recently launched an additive for urine (Streck, 2017). To date, there are no independent comparisons between Streck tube urine preservation methods and EDTA addition for the analysis of urinary mutDNA yield and would be of great importance to allow wider study participation.

### 5.1.2 Urinary DNA size

My findings suggest that longer fragments were not abundant in urinary cfDNA, particularly in whole urine, where I expected to see a greater proportion of cellular DNA released during lysis stage of DNA extraction, were initially surprising. A potential reason for this may have been that 360bp amplicon would be expected to give a lower count than 97bp fragments (or shorter) due to the preferential amplification of shorter fragments and that primers targeting a 97bp apmlicon are more likely to find primer landing and take off sites on a DNA fragment than primers producing a longer as the longer DNA region has more chance of being fragmented within the target amplicon. A potential control for this could have been to use size specific controls with un-fragmented DNA. However, during the course of studies, several key publications corroborated our initial results. In 2012, Tsui *et al.* described the size profile of urinary cfDNA extracted from spun urine (Tsui *et al.*, 2012). Of note, Tsui *et al.* demonstrated a 29bp peak in adult cfDNA, using WGS. Potential biases may occur at every step of cfDNA analysis. Briefly, shorter fragments may be preferentially extracted over fragments several hundreds of bps in length, when using column based extraction kits. Secondly library preparation methods may preferentially amplify shorter DNA fragments. Lastly, the NGS chemistry used by Tsui *et al.* and to generate these results, select for shorter fragments.

By using sWGS data of USN samples from our MIBC study, plots of DNA fragments size versus frequency were created, similar to those of Tsui *et al.* (Tsui *et al.*, 2012). Figure 5.1 shows urinary cfDNA size distribution for a healthy volunteer who contributed urine to the urine optimisation study, see section 2.4.1. The distribution of fragment sizes appears to match that of maternal urine described by Tsui *et al.* (Tsui *et al.*, 2012).

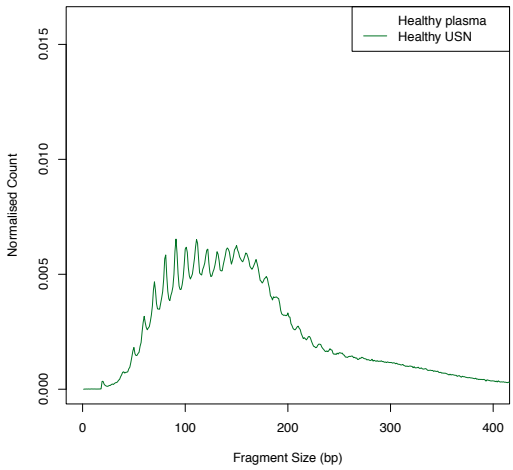



**Figure 5.1: DNA fragment size distribution for healthy urine.** Fragment sizes (x axis) are compared to their frequency (y axis).

When considering urinary cfDNA sizes from samples with detectable CNAs, a noticeable shift towards shorter size distributions is observed, see Figure 5.2. Indeed, the distribution peak shifts from 144bp in healthy and CNN urinary cfDNA samples to 100bp for mutDNA samples.

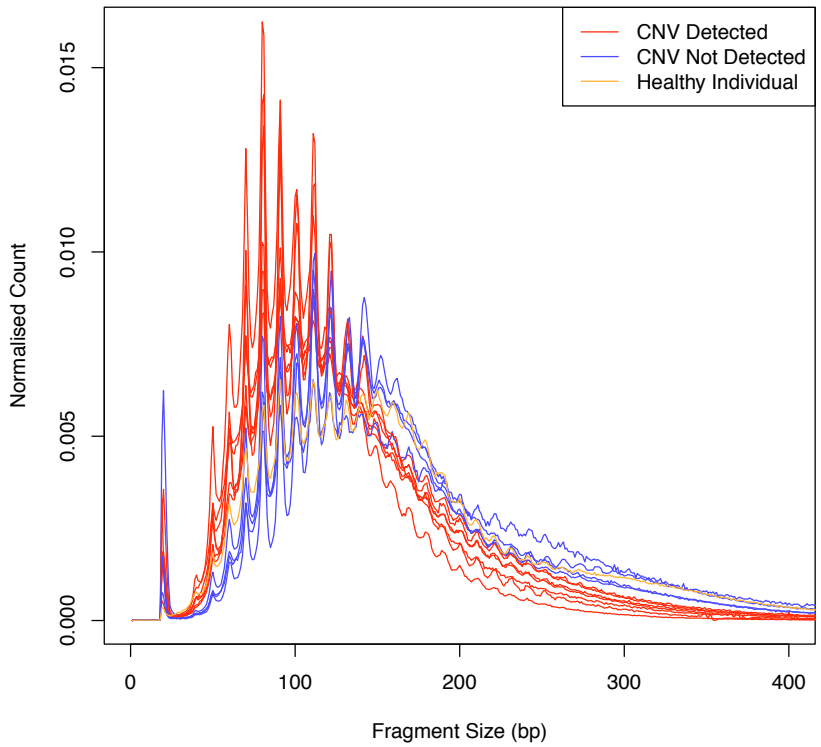



**Figure 5.2: Fragment size distribution showing the differing sizes of urinary cfDNA with detectable CNAs, urinary cfDNA without detectable CNAs and from healthy volunteers.**

Of interest is our finding that mutDNA is more often found in shorter fragments of urinary cfDNA. This extends previous work by Jiang *et al.* who found that CNAs were preferentially found in shorter plasma cfDNA fragments in 90 patients with hepatocellular

carcinoma, (Jiang *et al.*, 2015). Together, these results may help improve the understanding mutDNA generation and may help design more efficient mutDNA analysis methods (Mouliere and Rosenfeld, 2015).

### 5.1.3 Detection of mutDNA for urine processing study

The inability to detect mutDNA in the urine processing arm of the study is unfortunate and is likely to be multifactorial. Firstly only samples from 2 patients were obtained limiting the ability of our panel to detect mutations. Secondly, the panel used only detects mutations in 72% of cancers (Table 3.1) and a more broad panel design may have been able to detect SNVs in these samples. Specifically, the inability to target *TERT* will have affected the ability to detect SNVs in more cases. Indeed, *TERT* has been shown to harbour promoter region SNVs in 66% of patients with bladder cancer (Cancer Genome Atlas Research, 2014; Forbes *et al.*). These promoter SNVs occur within a highly repetitive region of the human genome. As cfDNA is likely to be short (~166bp), short amplicons would need to be designed to target *TERT* mutations and would therefore fall within the repetitive region, losing specificity. Indeed, where non-invasive *TERT* detection has been reported, it has been in more intact DNA e.g. urine cell pellet (Kinde *et al.*, 2013b; Fedriga *et al.*, 2001; Allory *et al.*, 2014; Ward *et al.*, 2016), allowing targeting with longer amplicons than would be suitable for cfDNA detection. As all other primers worked and >70% of SNVs would be captured by the existing panel no further attempts at targeting *TERT* were made. Lastly, the TUR samples for these patients were unavailable. Analysis of the TUR samples may have allowed identification of SNVs that could then have been tracked in detection mode or alternatively, exome sequencing of the tumour could have been performed and assay design to patient specific mutations.

## 5.2 MutDNA in MIBC

MutDNA analysis has been shown to have translational potential in many solid cancers (Bettegowda *et al.*, 2014). MutDNA has been studied in plasma but little data has been presented on the analysis of mutDNA in urine. In our cohort of MIBC patients we confirm the presence of SNVs and CNAs previously reported in BC (Cancer Genome Atlas Research; Forbes *et al.*). We confirm that mutations found in TUR samples are detectable in both the plasma and urine of BC patients. Our data, based upon the analysis of 86 time-points, demonstrate that mutDNA is detected more frequently and at higher levels in urine (both USN and UCP) and agrees with previous work (Birkenkamp-Demtroder *et al.*).

Importantly however, in some cases mutDNA was only detected in one individual sample type (including plasma). These private mutations could represent local anatomy or biology. For example, one could speculate that mutations private to plasma represent tumour clones situated deeper in muscle and/or closer to the vasculature. This analysis may be important as such clones could have different clinical outcomes. In this study, the observation of such plasma-specific mutations did not associate with response to NAC or early recurrence, though the number of patients was small. Nonetheless, our data indicates an advantage in assessing both urine (UCP & USN) and plasma for the comprehensive analysis of mutDNA in MIBC. Additionally, the observation that urinary AFs were higher than plasma AFs, suggests that the presence of mutDNA in urinary samples is due to direct tumour shedding. Beyond BC, this study confirms the role of sampling peripheral fluids in close proximity to diseased organs in order to improve the sensitivity of the detection of mutDNA (Wang *et al.;* Patel and Tsui).

Of potential clinical significance, we found that detection of mutDNA at the second cycle of NAC (2-3 weeks after initiation of therapy, depending on the specific regimen) using our methods was predictive of early disease recurrence, with a sensitivity of 83% and specificity of 100%. We acknowledge that we have studied a small number of cases and with a relatively short follow-up period (possibly resulting in recurrence events being missed). However, it is noteworthy that MIBC often recurs within 2 years (Zehnder *et al.*) and our recurrence rates are comparable with published data (Witjes *et al.*). Our proof of principle study therefore encourages further large-scale investigation. In our series, the single false negative case was likely due to the narrow focus of our bladder-specific panel, since we only detected a *PIK3CA* mutation in the TUR sample at low AF, which likely represents a minor sub-clone (similar to those described in patients 12 and 15). Analysis of additional targets, through the use of an expanded panel for targeted sequencing or by capture-based strategies (as has been used in other cancer types (Murtaza *et al.;* Newman *et al.*)), may have identified alternative cancer pathway mutations in this patient. Whilst we attempted to overcome this by interrogating the wider cancer genome through CNA profiling (Cancer Genome Atlas Research), we note that CNA analysis is limited to detecting mutDNA at ~5% mutant:wild type AF ratio (Heitzer *et al.*). Alternatively, by applying only *TP53* detection to the USN of the 9 patients who had *TP53* mutations in their TUR, a focussed *TP53* urinary assay would have detected all 4 patients that recurred (Figure 3.11).

If confirmed by larger studies, mutDNA detection in pre-NAC samples and in samples taken before the 2nd NAC cycle may be used to stratify patients into 3 groups (Figure 5.3). Firstly, those that are negative at both time-points may have a low (or no) burden of disease. Secondly, those that have a positive pre-NAC sample but negative samples at the 2nd NAC cycle are benefiting from their NAC and subsequent definitive therapy. The final group of patients has detectable mutDNA at both time-points and often progresses. Patients in this group are not benefiting from NAC and should be considered for expedited definitive therapy or alternative systemic therapy (e.g. targeted therapy or immune checkpoint inhibition). The specific mutations arising (or persisting) during therapy may inform further therapeutic strategy. Analysis of the presence of mutDNA in samples from later time-points was also informative but, would be less clinically useful. An early decision on continuation of NAC would prevent administration of multiple cycles of toxic therapy that is not benefiting the patient (Figure 3.5B).

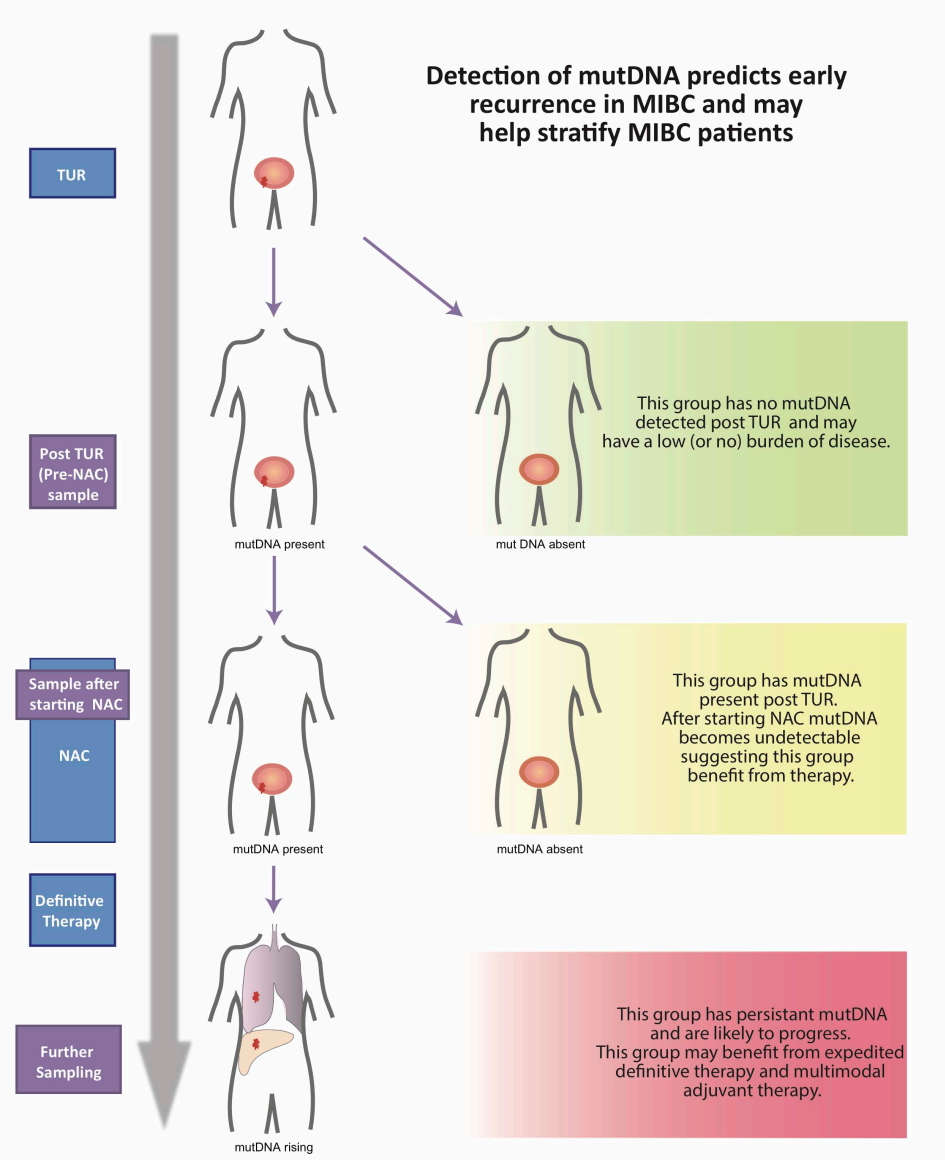



**Figure 5.3: MutDNA analysis may help stratify patients with MIBC in the future.** Based on our data, we have generated a model by which a patient's outcome can be predicted by mutDNA status before and during NAC. Specifically, mutDNA status may be used to stratify patients into 3 groups. Firstly, patients with undetectable mutDNA pre-NAC and during NAC, have a low (or no) burden of disease (green). Secondly, patients with detected mutDNA pre-NAC but with negative samples at the 2nd NAC cycle are likely to benefit from their NAC and subsequent definitive therapy (yellow). Finally, patients with mutDNA detected before and during NAC are unlikely to benefit from continued NAC and often progress (red). These patients should be considered for expedited definitive treatment or alternative treatments (e.g. targeted therapy or immune checkpoint inhibition).

These results represent the first description of tumour evolution through *de novo* mutation detection in urinary specimens. We hypothesise that the changes in the mutational profiles observed in these patients represent tumour evolution under the selective forces of surgery and NAC. This is exemplified by patient 15. It is likely that multiple clones were present initially, with analysis of initial urinary samples highlighting a dominant somatic profile consisting of *TP53* H193A and *YAP1* gain. Initiation of NAC allowed for the apparent emergence of a distinct (presumably NAC resistant) tumour clone, containing, *TP53* R273C, *NFE2L2* G31A and *CDKN2A* loss. Meanwhile the clone containing *TP53*

H193A and *YAP1* gain mutations appears to respond well to NAC and recedes. Our data emphasises the presence of multiple concomitant tumour clones and, the importance of monitoring mutDNA with an unbiased sequencing panel (e.g. one targeting all of *TP53*), rather than relying solely on the detection of known mutations from matched tumour samples. In addition, though we cannot rule out the effects of tumour heterogeneity in biasing our radical cystectomy sample analysis, the equivalent longitudinal data from peripheral fluid samples suggests that a combined body fluid sampling method is capable of overcoming the effects of spatially distinct clones. Overall, the data demonstrates the strength and potential of mutDNA profiling for non-invasive monitoring of clonal dynamics throughout therapy.

This proof of principle study has several weaknesses. Patient numbers were relatively small and thus firm clinical conclusions cannot be made. However, by studying several biological samples per patient at various time-points during NAC, the results establish a comprehensive overview of mutDNA in BC and generate specific hypotheses that can now be tested in prospective, sufficiently-powered, clinical studies. In addition, although the presence of mutDNA at the second cycle of NAC was associated with clinical outcome, we found instances where patients with favorable outcome had mutDNA detected in samples at later time-points. At this point, it is unclear whether these findings are clinically meaningful. However, it is clear that patients with a favorable outcome tend to lose detectable presence of mutDNA (Figure 3.5B). Finally, the detection methods chosen have limitations in panel size, sequence depth and sensitivity. These could be supplemented with the inclusion of other genomic regions of interest, e.g. *ERBB2, ERCC2* or other DNA damage pathway genes, mutations in which have been shown to predict responsiveness to NAC (Van Allen *et al.*, 2014; Groenendijk *et al.*, 2016; Yap *et al.*, 2014). However, the current panel design represents a reasonable trade-off in comparison with methods like digital PCR (high coverage but few mutations analysed) and capture-based NGS assays (large gene panels but coverage limited by sequencing costs).

Together with previous work, this study demonstrates the utility of urinary mutDNA analysis across the full spectrum of bladder cancer. Using multiple sample sources and complementary genetic techniques we have provided a more rigorous analysis of mutDNA. Furthermore, despite the small numbers of patients, this prospective study highlights the important role that mutDNA analysis can have in predicting aggressive disease in MIBC and could offer an opportunity for patients to consider expedited definitive therapy or alternative regimens.

## 5.3 ctDNA in Localised Prostate cancer

Our attempts to use a targeted panel to detect mutations in localised prostate cancer were unsuccessful. This is likely due to several factors. Firstly, the low mutation rate of prostate cancer, reportedly <0.9 mutations /Mb (Berger *et al.*, 2011) in localised disease and even as low as 2 mutations /Mb in heavily treated CRPC (Grasso *et al.*, 2012) suggests that there will be few mutations to track. Secondly, though mutations may be present, there are few recurrent mutations (Grasso *et al.*, 2012; Barbieri and Tomlins, 2014) making it difficult to design prostate cancer specific panels. Lastly, the multi-focality of prostate cancer (Cooper *et al.*, 2015), will mean that identifying clones at a localised disease stage to track using personalised panels will also be problematic. These factors are likely to result in prostate cancers being one of the most challenging solid cancers to monitor at the localised stage.

However, despite the multi-focality of localised prostate cancer, Lui *et al.* have demonstrated that most metastatic prostate cancers have a monoclonal origin through SNP and copy number analysis (Liu *et al.*, 2009). More recently, others have confirmed this through exome sequencing of CRPC and matched localised prostate cancer samples (Grasso *et al.*, 2012; Carreira *et al.*, 2014). If the potentially malignant clone can be detected at an early stage, ctDNA analysis may help identify men with poor prognoses. There remain three important challenges to overcome before such an approach could be viable. Firstly, whether recurrent mutations are present in CRPC. Secondly, whether clones containing these mutations are present in plasma at localised prostate cancer stage. Thirdly, whether individual clones are detectable in amongst multifocal and heterogeneous localised prostate cancer (Andreoiu and Cheng, 2010; Cooper *et al.*, 2015).

### 5.3.1 *TP53* as a marker of aggressive disease

Several studies suggest that CRPC may have recurrent mutations and in particular may harbour *TP53* SNVs. Indeed, out of 50 CRPC cases investigated by Grasso *et al.*, 21 cases had *TP53* SNVs (including 1 case with a 2 base deletion) (Grasso *et al.*, 2012), whilst Beltran *et al.* found *TP53* mutations in 40% of cases (Beltran *et al.*, 2013). In our study of 19 metastatic samples from men with CRPC, 11 *TP53* SNVs were detected, confirming the association of *TP53* with CRPC. Furthermore, in all men who had prostatectomy samples available for analysis (n=6), the same *TP53* SNV present in the

dominant malignant clone was found. Though the numbers studied are small, these results lend weight to the hypothesis that lethal CRPC may have monoclonal origin (Grasso *et al.*, 2012; Liu *et al.*, 2009) and confirms Haffner *et al.'s* observation that the dominant clone during malignancy can be traced back to a clone present at the localised prostate cancer stage (Haffner *et al.*, 2013).

In addition, historical pre-operative plasma samples in 3 men were subjected to amplicon sequencing to identify the same *TP53* SNVs that would dominate future malignant clones, in 2/3 of the men's samples. These results build on Carreira *et al.'s* findings that 3/16 of men *TP53* SNVs, detectable in pre-castration resistant disease persist at the CRPC stage, though the pre-castration stage is likely metastatic due to the high AFs, (Carreira *et al.*, 2014). Our proof of principle finding is an important contribution to knowledge and has significant clinical implications. Though the numbers are small, the findings suggest that acquisition of *TP53* SNVs may confer metastatic potential and that this aggressive change can be detected in plasma at the localised stage of disease. If corroborated in larger studies, ctDNA analysis could help identify potentially aggressive prostate cancers or help monitor prostate cancer during active surveillance, through timely detection of aggressive changes. These results warrant further large-scale trial in longitudinal studies.

### 5.3.2 Accounting for prostate heterogeneity

When considering the clinical application of a ctDNA test, the multi-focality and heterogeneity of prostate cancer will impact on the ability to track aggressive clones in localised disease. Indeed, Figure 4.15 demonstrates that current standards of amplicon sequencing alone will not allow the identification of ctDNA from clonally heterogeneous primary prostate tumours. MRD-Seq was developed and used to demonstrate the presence of mutations in plasma samples taken at the localised stage, representing discrete tumour regions from multifocal disease.

Dilution experiments demonstrate the analytical sensitivity of MRD-Seq to detect AFs down to 0.09% (the limit of sensitivity) and suggests a background error rate of 0.02% AF (MCDNA), see Table 4.19. Figure 4.27 suggests that the actual limit of sensitivity for MRD-Seq could be improved with further dilutions, thus pushing the actual limit of sensitivity closer to the background error rate. Thus the ability of MRD-Seq to assess multiple amplicons is likely to be between 1 mutant in 1,100 – 5,000 mutant:wild-type copies. This improves on the previously reported analytical sensitivity of Tam-Seq in detection mode of 1 in 700 mutant:wild-type molecules (Forshew *et al.*, 2012) and would be similar to

standard dPCR techniques (Vogelstein and Kinzler, 1999). MRD-Seq is however, likely to be less sensitive than more advanced dPCR techniques, e.g. droplet dPCR, with a reported sensitivity of detecting 1 in 13,000 mutant:wild-types (Milbury *et al.*, 2014), although such sensitivities can rarely be achieved in the analysis of cfDNA where starting material is frequently limited to ~$10^3$ copies/ml plasma.

MRD-Seq does however, have the added benefit of tracking multiple mutations, reducing the risk of missing disease recurrence through clonal evolution when applied to longitudinal sample collections. Developing the idea of tracking multiple mutations, Newman *et al.* described Capp-Seq, which leverages thousands of mutations, to obtain a reported analytical sensitivity to detect 1 in 5000 mutant:wild-type molecules (Newman *et al.*, 2014). Further computational refinements have improved the sensitivity to detect 1 in 25,000 mutant:wild-type molecules (Newman *et al.*, 2016) outperforming MRD-Seq for analytical sensitivity by ~5 fold when input material is not a limiting factor.

However, Capp-Seq is likely to be more costly than MRD-Seq as it relies on DNA capture technology. With the majority of the cost per sample due to the expense of capture baits and the need for deep sequencing of large regions of the genome. Conversely, the current cost of MRD-Seq for assessing a single plasma sample using 50 replicates is approximately £650, though this excludes the costs of exome sequencing / WGS of the tumour, and purchasing the primer panel, both of which are one off costs. Furthermore, it is likely that the cost of MRD-Seq can be significantly reduced through optimisation. In particular, investigating the barcoding of individual replicates during pre-amplification to enable pooling the replicates prior to library size selection, could significantly lower the cost. This may be achieved by using molecular barcodes (Peng *et al.*, 2015) for each replicate, though the cost of primers may increase, or by optimising PCR conditions to omit the microfluidic singleplex PCR step. These improvements could conceivably bring the cost of MRD-Seq down to £300-400 per sample and within the reach of current recurrence detection tests, for example, a single flexible cystoscopy attendance costs ~£430, whilst a CT scan costs ~£150 (Monitor, 2016). Therefore, MRD-Seq may be a cost effective method to longitudinally screen plasma samples for minimal residual disease or recurrence. This is likely to be useful in cancer types where the detection of minimal residual disease is an area of need e.g. breast cancer, cervical cancer, glioma, etc., though for prostate cancer, PSA detection is a suitable, cheaper alternative.

The limited coverage obtained from the pre-operative plasma samples, limits the ability to draw firm conclusions about whether clonal heterogeneity can be monitored in localised prostate cancer. This aspect could not be further evaluated due to the lack of sample material. That the pre-operative plasma DNA alone suffered from poor coverage, whilst tumour dilutions and healthy plasma control coverage was as expected, suggest the possibility of technical issues with the pre-operative plasma samples. Possible explanations include poor initial processing (low yields or contaminant carry over), and the age of plasma samples (collected in December 2012). Indeed, Sozzi *et al.* applied quantitative PCR of *hTERT* to paired lung cancer plasma samples to show that the quantity of DNA isolated decreases by approximately 30% per year despite storage at -80°C (Sozzi *et al.*, 2005). If this holds true for the prostate plasma samples, after 5 years of storage the quantity of cfDNA may be ~16.8% of the material originally collected.

In addition, the ECDF distribution of AFs for pre-operative plasma and HVP samples were not statistically significantly different. This could be due to the use of different sequencing platforms, the lower number of HVP samples used and the differing coverage achieved by MRD-Seq on plasma and HVP samples (as every base will have a unique non-reference error profile). In addition, the global signal may be dominated by technical noise and using the detection threshold (1 / number of copies per replicate) leverages the potential benefits of MRD-Seq. Analysis of pre-operative plasma DNA using this detection threshold approach suggests the presence of private mutations from clonally distinct regions of prostate cancer in plasma DNA. In our proof-of-principle study, some clones were better represented in the plasma DNA e.g. (regions T1 and T2 for man 7), whilst others were less well represented e.g. (region T5 from man 7). The reasons for this are at present poorly understood but could include features of the clones in these regions; proximity to tumour vasculature, size, grading and cell turnover rate. These factors together with the clinical relevance of differing clonal representation in the plasma need further investigation.

These proof-of-principle experiments demonstrate that ctDNA analysis of pre-operative plasma samples may allow the detection of potentially aggressive *TP53* SNVs and may demonstrate the heterogeneity of localised prostate cancer. Taken together ctDNA analysis may allow the detection potentially aggressive mutational changes in a clone at an early stage and warrants further investigation. Further work is needed to optimise the MRD-Seq approach and to test these preliminary findings in larger patient cohorts, where detecting minimal residual disease is an unmet clinical need.

## 5.4 Future Work

Work presented in this thesis represent important contribution to knowledge and some aspects have the potential for translation into clinical practice and warrant further investigation. For bladder cancer, the urinary processing for the detection of mutDNA could be further refined. This could take the form of increasing the number of commercially available kits being assessed and increasing the number of samples used or introducing more nuanced processing steps e.g., comparing 0mM, 5mM and 10mM concentrations of EDTA addition. More importantly however, would be an assessment of whether urine processing and DNA extraction is optimised for mutDNA detection. Specifically this should involve a rigorous assessment of whether mutDNA is more prominent in whole urine samples, or urine supernatant. In MIBC initial results seem encouraging that mutDNA analysis may detect early treatment failure, so that further adjuvant therapies may be used. In the first instance, these results would need corroboration in larger cohorts of patients ideally across multiple institutions.

In prostate cancer, though we (Hong *et al.*, 2015) and others (Kumar *et al.*, 2011; Beltran *et al.*, 2013) have identified some potentially aggressive drivers, a complete understanding of events and their timings remain elusive and requires further investigation through multi-spatial genome/exome wide sequencing of targeted populations of men and functional studies. The finding that potentially malignant clones are detectable in the plasma of men at the stage of localised disease warrants validation in large-scale studies, especially considering the scope of potential benefit. Furthermore, our findings demonstrate that the analysis of cfDNA, based on the MRD-Seq principles of tracking multiple targets and using diluted replicates, may be sensitive enough to monitor ctDNA in localised prostate cancer. If corroborated in larger cohorts, these methods can be adapted to assay potentially aggressive markers of prostate cancer. Indeed, the refinement of ultra-sensitive ctDNA detection techniques and the known markers of aggression are required before implementation of a ctDNA test for aggressive prostate cancer can be considered.

# REFERENCES


ALLORY, Y., BEUKERS, W., SAGRERA, A., FLANDEZ, M., MARQUES, M., MARQUEZ, M., VAN DER KEUR, K. A., DYRSKJOT, L., LURKIN, I., VERMEIJ, M., CARRATO, A., LLORETA, J., LORENTE, J. A., CARRILLO-DE SANTA PAU, E., MASIUS, R. G., KOGEVINAS, M., STEYERBERG, E. W., VAN TILBORG, A. A., ABAS, C., ORNTOFT, T. F., ZUIVERLOON, T. C., MALATS, N., ZWARTHOFF, E. C. & REAL, F. X. 2014. Telomerase reverse transcriptase promoter mutations in bladder cancer: high frequency across stages, detection in urine, and lack of association with outcome. *Eur Urol,* 65**,** 360-6.

ANDREOIU, M. & CHENG, L. 2010. Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. *Hum Pathol,* 41**,** 781-93.

ANDRIOLE, G. L., CRAWFORD, E. D., GRUBB, R. L., 3RD, BUYS, S. S., CHIA, D., CHURCH, T. R., FOUAD, M. N., ISAACS, C., KVALE, P. A., REDING, D. J., WEISSFELD, J. L., YOKOCHI, L. A., O'BRIEN, B., RAGARD, L. R., CLAPP, J. D., RATHMELL, J. M., RILEY, T. L., HSING, A. W., IZMIRLIAN, G., PINSKY, P. F., KRAMER, B. S., MILLER, A. B., GOHAGAN, J. K., PROROK, P. C. & TEAM, P. P. 2012. Prostate cancer screening in the randomized Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial: mortality results after 13 years of follow-up. *J Natl Cancer Inst,* 104**,** 125-32.

ANTONATOS, D., PATSILINAKOS, S., SPANODIMOS, S., KORKONIKITAS, P. & TSIGAS, D. 2006. Cell-Free DNA Levels as a Prognostic Marker in Acute Myocardial Infarction. *Annals of the New York Academy of Sciences,* 1075**,** 278-281.

BABJUK, M., BÖHLE, A., BURGER, M., COMPÉRAT, E., KAASINEN, E., PALOU, J., VAN RHIJN, B., ROUPRÊT, M., SHARIAT, S., SYLVESTER, R. & ZIGEUNER, R. 2014. Guidelines on Non-muscle-invasive Bladder Cancer (Ta, T1 and CIS). European Association of Urology.

BARBIERI, C. E., BACA, S. C., LAWRENCE, M. S., DEMICHELIS, F., BLATTNER, M., THEURILLAT, J. P., WHITE, T. A., STOJANOV, P., VAN ALLEN, E., STRANSKY, N., NICKERSON, E., CHAE, S. S., BOYSEN, G., AUCLAIR, D., ONOFRIO, R. C., PARK, K., KITABAYASHI, N., MACDONALD, T. Y., SHEIKH, K., VUONG, T., GUIDUCCI, C., CIBULSKIS, K., SIVACHENKO, A., CARTER, S. L., SAKSENA, G., VOET, D., HUSSAIN, W. M., RAMOS, A. H., WINCKLER, W., REDMAN, M. C., ARDLIE, K., TEWARI, A. K., MOSQUERA, J. M., RUPP, N., WILD, P. J., MOCH, H., MORRISSEY, C., NELSON, P. S., KANTOFF, P. W., GABRIEL, S. B., GOLUB, T. R., MEYERSON, M., LANDER, E. S., GETZ, G., RUBIN, M. A. & GARRAWAY, L. A. 2012. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet,* 44**,** 685-9.

BARBIERI, C. E. & TOMLINS, S. A. 2014. The prostate cancer genome: Perspectives and potential. *Urologic Oncology: Seminars and Original Investigations,* 32**,** 53.e15-53.e22.

BARRA, G. B., SANTA RITA, T. H., DE ALMEIDA VASQUES, J., CHIANCA, C. F., NERY, L. F. & SANTANA SOARES COSTA, S. 2015. EDTA-mediated inhibition of DNases protects circulating cell-free DNA from ex vivo degradation in blood samples. *Clin Biochem,* 48**,** 976-81.

BARTOLONI, E., LUDOVINI, V., ALUNNO, A., PISTOLA, L., BISTONI, O., CRINO, L. & GERLI, R. 2011. Increased levels of circulating DNA in patients with systemic autoimmune diseases: A possible marker of disease activity in Sjogren's syndrome. *Lupus,* 20**,** 928-35.

BEASLEY, M., WILLIAMS, S. G., PICKLES, T. & UNIT, B. P. O. 2008. Expanded risk groups help determine which prostate radiotherapy sub-group may benefit from adjuvant androgen deprivation therapy. *Radiat Oncol,* 3**,** 8.

BEAVER, J. A., JELOVAC, D., BALUKRISHNA, S., COCHRAN, R. L., CROESSMANN, S., ZABRANSKY, D. J., WONG, H. Y., TORO, P. V., CIDADO, J., BLAIR, B. G., CHU, D., BURNS, T., HIGGINS, M. J., STEARNS, V., JACOBS, L., HABIBI, M., LANGE, J., HURLEY, P. J., LAURING, J., VANDENBERG, D. A., KESSLER, J., JETER, S., SAMUELS, M. L., MAAR, D., COPE, L., CIMINO-MATHEWS, A., ARGANI, P., WOLFF, A. C. & PARK, B. H. 2014. Detection of Cancer DNA in Plasma of Patients with Early-Stage Breast Cancer. *Clinical Cancer Research,* 20**,** 2643-2650.

BELTRAN, H., YELENSKY, R., FRAMPTON, G. M., PARK, K., DOWNING, S. R., MACDONALD, T. Y., JAROSZ, M., LIPSON, D., TAGAWA, S. T., NANUS, D. M., STEPHENS, P. J., MOSQUERA, J. M., CRONIN, M. T. & RUBIN,

M. A. 2013. Targeted Next-generation Sequencing of Advanced Prostate Cancer Identifies Potential Therapeutic Targets and Disease Heterogeneity. *Eur Urol,* 63**,** 920-6.

BERGER, M. F., LAWRENCE, M. S., DEMICHELIS, F., DRIER, Y., CIBULSKIS, K., SIVACHENKO, A. Y., SBONER, A., ESGUEVA, R., PFLUEGER, D., SOUGNEZ, C., ONOFRIO, R., CARTER, S. L., PARK, K., HABEGGER, L., AMBROGIO, L., FENNELL, T., PARKIN, M., SAKSENA, G., VOET, D., RAMOS, A. H., PUGH, T. J., WILKINSON, J., FISHER, S., WINCKLER, W., MAHAN, S., ARDLIE, K., BALDWIN, J., SIMONS, J. W., KITABAYASHI, N., MACDONALD, T. Y., KANTOFF, P. W., CHIN, L., GABRIEL, S. B., GERSTEIN, M. B., GOLUB, T. R., MEYERSON, M., TEWARI, A., LANDER, E. S., GETZ, G., RUBIN, M. A. & GARRAWAY, L. A. 2011. The genomic complexity of primary human prostate cancer. *Nature,* 470**,** 214-20.

BETTEGOWDA, C., SAUSEN, M., LEARY, R. J., KINDE, I., WANG, Y. X., AGRAWAL, N., BARTLETT, B. R., WANG, H., LUBER, B., ALANI, R. M., ANTONARAKIS, E. S., AZAD, N. S., BARDELLI, A., BREM, H., CAMERON, J. L., LEE, C. C., FECHER, L. A., GALLIA, G. L., GIBBS, P., LE, D., GIUNTOLI, R. L., GOGGINS, M., HOGARTY, M. D., HOLDHOFF, M., HONG, S. M., JIAO, Y. C., JUHL, H. H., KIM, J. J., SIRAVEGNA, G., LAHERU, D. A., LAURICELLA, C., LIM, M., LIPSON, E. J., MARIE, S. K. N., NETTO, G. J., OLINER, K. S., OLIVI, A., OLSSON, L., RIGGINS, G. J., SARTORE-BIANCHI, A., SCHMIDT, K., SHIH, I. M., OBA-SHINJO, S. M., SIENA, S., THEODORESCU, D., TIE, J. N., HARKINS, T. T., VERONESE, S., WANG, T. L., WEINGART, J. D., WOLFGANG, C. L., WOOD, L. D., XING, D. M., HRUBAN, R. H., WU, J., ALLEN, P. J., SCHMIDT, C. M., CHOTI, M. A., VELCULESCU, V. E., KINZLER, K. W., VOGELSTEIN, B., PAPADOPOULOS, N. & LUIS, A. J. 2014. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Science Translational Medicine,* 6**,** 224ra24.

BIASSONI, R. & RASO, A. 2014. *Quantitative Real-Time PCR*, Humana Press.

BIRKENKAMP-DEMTRODER, K., NORDENTOFT, I., CHRISTENSEN, E., HOYER, S., REINERT, T., VANG, S., BORRE, M., AGERBAEK, M., JENSEN, J. B., ORNTOFT, T. F. & DYRSKJOT, L. 2016. Genomic Alterations in Liquid Biopsies from Patients with Bladder Cancer. *Eur Urol,* 70**,** 75-82.

BOTEZATU, I., SERDYUK, O., POTAPOVA, G., SHELEPOV, V., ALECHINA, R., MOLYAKA, Y., ANANEV, V., BAZIN, I., GARIN, A., NARIMANOV, M., KNYSH, V., MELKONYAN, H., UMANSKY, S. & LICHTENSTEIN, A. 2000. Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism. *Clin Chem,* 46**,** 1078-84.

BOTTEMAN, M. F., PASHOS, C. L., REDAELLI, A., LASKIN, B. & HAUSER, R. 2003. The health economics of bladder cancer: a comprehensive review of the published literature. *Pharmacoeconomics,* 21**,** 1315-30.

BOUSTEAD, G. B., FOWLER, S., SWAMY, R., KOCKLEBERGH, R., HOUNSOME, L. & SECTION OF ONCOLOGY, B. 2014. Stage, grade and pathological characteristics of bladder cancer in the UK: British Association of Urological Surgeons (BAUS) urological tumour registry. *BJU Int,* 113**,** 924-30.

BOYD, L. K., MAO, X. & LU, Y. J. 2012. The complexity of prostate cancer: genomic alterations and heterogeneity. *Nat Rev Urol,* 9**,** 652-64.

BRANDAU, S. & SUTTMANN, H. 2007. Thirty years of BCG immunotherapy for non-muscle invasive bladder cancer: a success story with room for improvement. *Biomed Pharmacother,* 61**,** 299-305.

BRONKHORST, A. J., AUCAMP, J. & PRETORIUS, P. J. 2015. Cell-free DNA: Preanalytical variables. *Clin Chim Acta,* 450**,** 243-53.

BUSTIN, S. 2010. *The PCR Revolution BASIC TECHNOLOGIES AND APPLICATIONS,* New York, CAMBRIDGE UNIVERSITY PRESS.

CANCER GENOME ATLAS RESEARCH, N. 2014. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature,* 507**,** 315-22.

CANCER GENOME ATLAS RESEARCH, N. 2015. The Molecular Taxonomy of Primary Prostate Cancer. *Cell,* 163**,** 1011-25.

CARREIRA, S., ROMANEL, A., GOODALL, J., GRIST, E., FERRALDESCHI, R., MIRANDA, S., PRANDI, D., LORENTE, D., FRENEL, J.-S., PEZARO, C., OMLIN, A., RODRIGUES, D. N., FLOHR, P., TUNARIU, N., S. DE BONO, J., DEMICHELIS, F. & ATTARD, G. 2014. Tumor clone dynamics in lethal prostate cancer. *Science Translational Medicine,* 6**,** 254ra125.

CERAMI, E., GAO, J., DOGRUSOZ, U., GROSS, B. E., SUMER, S. O., AKSOY, B. A., JACOBSEN, A., BYRNE, C. J., HEUER, M. L., LARSSON, E., ANTIPIN, Y., REVA, B., GOLDBERG, A. P., SANDER, C. & SCHULTZ, N. 2012. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov,* 2**,** 401-4.

CHEN, D., PAN, S., ZHANG, S., HUANG, P., XIA, W., XIE, E., GU, B., WANG, F., XU, J., XU, T., LU, Y., YANG, D. & LU, S. 2011. The Clinical Significance of Plasma DNA Quantification for Quake Trauma Patients. *In:* GAHAN, P. B. (ed.) *Circulating Nucleic Acids in Plasma and Serum.* Springer Netherlands.

CHEN, X., BONNEFOI, H., DIEBOLD-BERGER, S., LYAUTEY, J., LEDERREY, C., FALTIN-TRAUB, E., STROUN, M. & ANKER, P. 1999. Detecting tumor-related alterations in plasma or serum DNA of patients diagnosed with breast cancer. *Clin Cancer Res,* 5**,** 2297-303.

COOPER, C. S., EELES, R., WEDGE, D. C., VAN LOO, P., GUNDEM, G., ALEXANDROV, L. B., KREMEYER, B., BUTLER, A., LYNCH, A. G., CAMACHO, N., MASSIE, C. E., KAY, J., LUXTON, H. J., EDWARDS, S., KOTE-JARAI, Z., DENNIS, N., MERSON, S., LEONGAMORNLERT, D., ZAMORA, J., CORBISHLEY, C., THOMAS, S., NIK-ZAINAL, S., RAMAKRISHNA, M., O'MEARA, S., MATTHEWS, L., CLARK, J., HURST, R., MITHEN, R., BRISTOW, R. G., BOUTROS, P. C., FRASER, M., COOKE, S., RAINE, K., JONES, D., MENZIES, A., STEBBINGS, L., HINTON, J., TEAGUE, J., MCLAREN, S., MUDIE, L., HARDY, C., ANDERSON, E., JOSEPH, O., GOODY, V., ROBINSON, B., MADDISON, M., GAMBLE, S., GREENMAN, C., BERNEY, D., HAZELL, S., LIVNI, N., GROUP, I. P., FISHER, C., OGDEN, C., KUMAR, P., THOMPSON, A., WOODHOUSE, C., NICOL, D., MAYER, E., DUDDERIDGE, T., SHAH, N. C., GNANAPRAGASAM, V., VOET, T., CAMPBELL, P., FUTREAL, A., EASTON, D., WARREN, A. Y., FOSTER, C. S., STRATTON, M. R., WHITAKER, H. C., MCDERMOTT, U., BREWER, D. S. & NEAL, D. E. 2015. Analysis of the genetic phylogeny of multifocal prostate cancer identifies multiple independent clonal expansions in neoplastic and morphologically normal prostate tissue. *Nat Genet,* 47**,** 367-72.

COOPERBERG, M. R., BROERING, J. M. & CARROLL, P. R. 2009. Risk assessment for prostate cancer metastasis and mortality at the time of diagnosis. *J Natl Cancer Inst,* 101**,** 878-87.

CRUK. 2014. *Prostate cancer statistics* [Online]. Cancer Reaserch UK. Available: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/prostate/ [Accessed 04/09/2014].

CUZICK, J., BERNEY, D. M., FISHER, G., MESHER, D., MOLLER, H., REID, J. E., PERRY, M., PARK, J., YOUNUS, A., GUTIN, A., FOSTER, C. S., SCARDINO, P., LANCHBURY, J. S., STONE, S. & TRANSATLANTIC PROSTATE, G. 2012. Prognostic value of a cell cycle progression signature for prostate cancer death in a conservatively managed needle biopsy cohort. *Br J Cancer,* 106**,** 1095-9.

D'AMICO, A. V., WHITTINGTON, R., MALKOWICZ, S. B., SCHULTZ, D., BLANK, K., BRODERICK, G. A., TOMASZEWSKI, J. E., RENSHAW, A. A., KAPLAN, I., BEARD, C. J. & WEIN, A. 1998. Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer. *JAMA,* 280**,** 969-74.

DAWSON, S. J., TSUI, D. W., MURTAZA, M., BIGGS, H., RUEDA, O. M., CHIN, S. F., DUNNING, M. J., GALE, D., FORSHEW, T., MAHLER-ARAUJO, B., RAJAN, S., HUMPHRAY, S., BECQ, J., HALSALL, D., WALLIS, M., BENTLEY, D., CALDAS, C. & ROSENFELD, N. 2013. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med,* 368**,** 1199-209.

DE LA TAILLE, A., ANTIPHON, P., SALOMON, L., CHERFAN, M., PORCHER, R., HOZNEK, A., SAINT, F., VORDOS, D., CICCO, A., YIOU, R., ZAFRANI, E. S., CHOPIN, D. & ABBOU, C. C. 2003. Prospective evaluation of a 21-sample needle biopsy procedure designed to improve the prostate cancer detection rate. *Urology,* 61**,** 1181-6.

DENMEADE, S. R. & ISAACS, J. T. 2002. A history of prostate cancer treatment. *Nat Rev Cancer,* 2**,** 389-96.

DIAZ, L. A., JR., WILLIAMS, R. T., WU, J., KINDE, I., HECHT, J. R., BERLIN, J., ALLEN, B., BOZIC, I., REITER, J. G., NOWAK, M. A., KINZLER, K. W., OLINER, K. S. & VOGELSTEIN, B. 2012. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature,* 486**,** 537-40.

DIEHL, F., LI, M., DRESSMAN, D., HE, Y., SHEN, D., SZABO, S., DIAZ, L. A., JR., GOODMAN, S. N., DAVID, K. A., JUHL, H., KINZLER, K. W. & VOGELSTEIN, B. 2005. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc Natl Acad Sci U S A,* 102**,** 16368-73.

DIEHL, F., SCHMIDT, K., CHOTI, M. A., ROMANS, K., GOODMAN, S., LI, M., THORNTON, K., AGRAWAL, N., SOKOLL, L., SZABO, S. A., KINZLER, K. W., VOGELSTEIN, B. & DIAZ, L. A., JR. 2008. Circulating mutant DNA to assess tumor dynamics. *Nat Med,* 14**,** 985-90.

DIETRICH, H. & DIETRICH, B. 2001. Ludwig Rehn (1849-1930)--pioneering findings on the aetiology of bladder tumours. *World J Urol,* 19**,** 151-3.

DIEZ, D. M. 2013. OIsurv: Survival analysis supplement to OpenIntro guide. R package version 0.2 ed.

EICHLER, K., HEMPEL, S., WILBY, J., MYERS, L., BACHMANN, L. M. & KLEIJNEN, J. 2006. Diagnostic value of systematic biopsy methods in the investigation of prostate cancer: a systematic review. *J Urol,* 175**,** 1605-12.

EPSTEIN, J. I., ALLSBROOK, W. C., JR., AMIN, M. B., EGEVAD, L. L. & COMMITTEE, I. G. 2005. The 2005 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma. *Am J Surg Pathol,* 29**,** 1228-42.

EPSTEIN, J. I., EGEVAD, L., AMIN, M. B., DELAHUNT, B., SRIGLEY, J. R., HUMPHREY, P. A. & GRADING, C. 2016a. The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. *Am J Surg Pathol,* 40**,** 244-52.

EPSTEIN, J. I., ZELEFSKY, M. J., SJOBERG, D. D., NELSON, J. B., EGEVAD, L., MAGI-GALLUZZI, C., VICKERS, A. J., PARWANI, A. V., REUTER, V. E., FINE, S. W., EASTHAM, J. A., WIKLUND, P., HAN, M., REDDY, C. A., CIEZKI, J. P., NYBERG, T. & KLEIN, E. A. 2016b. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. *Eur Urol,* 69**,** 428-35.

FEDEWA, S. A., SOLIMAN, A. S., ISMAIL, K., HABLAS, A., SEIFELDIN, I. A., RAMADAN, M., OMAR, H. G., NRIAGU, J. & WILSON, M. L. 2009. Incidence analyses of bladder cancer in the Nile delta region of Egypt. *Cancer Epidemiol,* 33**,** 176-81.

FEDRIGA, R., GUNELLI, R., NANNI, O., BACCI, F., AMADORI, D. & CALISTRI, D. 2001. Telomerase activity detected by quantitative assay in bladder carcinoma and exfoliated cells in urine. *Neoplasia,* 3**,** 446-50.

FERLAY, J., SOERJOMATARAM, I., ERVIK, M., DIKSHIT, R., ESER, S., MATHERS, C., REBELO, M., PARKIN, D., FORMAN, D. & BRAY, F. 2013. *Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11* [Online]. Lyon, France: International Agency for Research on Cancer. Available: http://globocan.iarc.fr [Accessed 05/09/2014 2014].

FLUIDIGM. 2017. *Access Array User Guide_v4.pdf* [Online]. Available: https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwix_MvFwrbUAhWOYlAKHacnC60QFggmMAA&url=https%3A%2F%2Fwww.fluidigm.com%2Fbinaries%2Fcontent%2Fdocuments%2Ffluidigm%2Fresources%2Faa-illumina-100-3770%2Faa-illumina-100-3770%2Ffluidigm%3Afile&usg=AFQjCNFBLW4Jz7LqJmrKyQmDzAUV2lAz1A&sig2=NEKHrgu55Q7JN3_U1leXZA [Accessed 02/03/2017].

FORBES, S. A., BINDAL, N., BAMFORD, S., COLE, C., KOK, C. Y., BEARE, D., JIA, M., SHEPHERD, R., LEUNG, K., MENZIES, A., TEAGUE, J. W., CAMPBELL, P. J., STRATTON, M. R. & FUTREAL, P. A. 2011. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res,* 39**,** D945-50.

FORSHEW, T., MURTAZA, M., PARKINSON, C., GALE, D., TSUI, D. W. Y., KAPER, F., DAWSON, S.-J., PISKORZ, A. M., JIMENEZ-LINAN, M., BENTLEY, D., HADFIELD, J., MAY, A. P., CALDAS, C., BRENTON, J. D. & ROSENFELD, N. 2012. Noninvasive Identification and Monitoring of Cancer Mutations by Targeted Deep Sequencing of Plasma DNA. *Science Translational Medicine,* 4**,** 136ra68.

FOURNIÉ, G. J., COURTIN, J.-P., LAVAL, F., CHALÉ, J.-J., POURRAT, J. P., PUJAZON, M.-C., LAUQUE, D. & CARLES, P. 1995. Plasma DNA as a marker of cancerous cell death. Investigations in patients suffering from lung cancer and in nude mice bearing human tumours. *Cancer Letters,* 91**,** 221-227.

FREEDLAND, S. J., HUMPHREYS, E. B., MANGOLD, L. A., EISENBERGER, M., DOREY, F. J., WALSH, P. C. & PARTIN, A. W. 2005. Risk of prostate cancer-specific mortality following biochemical recurrence after radical prostatectomy. *JAMA,* 294**,** 433-9.

GAO, J., AKSOY, B. A., DOGRUSOZ, U., DRESDNER, G., GROSS, B., SUMER, S. O., SUN, Y., JACOBSEN, A., SINHA, R., LARSSON, E., CERAMI, E., SANDER, C. & SCHULTZ, N. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal,* 6**,** pl1.

GNANAPRAGASAM, V., LOPHATANANON, A., MUIR, K., GAVIN, A., WRIGHT, K. & GREENBERG, D. 2016. 613 An improved clinical risk stratification system to better predict cancer specific mortality at diagnosis in primary non-metastatic prostate cancer. *European Urology Supplements,* 15**,** e613.

GONZALEZ-MASIA, J. A., GARCIA-OLMO, D. & GARCIA-OLMO, D. C. 2013. Circulating nucleic acids in plasma and serum (CNAPS): applications in oncology. *Onco Targets Ther,* 6**,** 819-32.

GRASSO, C. S., WU, Y. M., ROBINSON, D. R., CAO, X., DHANASEKARAN, S. M., KHAN, A. P., QUIST, M. J., JING, X., LONIGRO, R. J., BRENNER, J. C., ASANGANI, I. A., ATEEQ, B., CHUN, S. Y., SIDDIQUI, J., SAM, L., ANSTETT, M., MEHRA, R., PRENSNER, J. R., PALANISAMY, N., RYSLIK, G. A., VANDIN, F., RAPHAEL, B. J., KUNJU, L. P., RHODES, D. R., PIENTA, K. J., CHINNAIYAN, A. M. & TOMLINS, S. A. 2012. The mutational landscape of lethal castration-resistant prostate cancer. *Nature,* 487**,** 239-43.

GREER, C. E., LUND, J. K. & MANOS, M. M. 1991. PCR amplification from paraffin-embedded tissues: recommendations on fixatives for long-term storage and prospective studies. *Genome Research,* 1**,** 46-50.

GROENENDIJK, F. H., DE JONG, J., FRANSEN VAN DE PUTTE, E. E., MICHAUT, M., SCHLICKER, A., PETERS, D., VELDS, A., NIEUWLAND, M., VAN DEN HEUVEL, M. M., KERKHOVEN, R. M., WESSELS, L. F., BROEKS, A., VAN RHIJN, B. W., BERNARDS, R. & VAN DER HEIJDEN, M. S. 2016. ERBB2 Mutations Characterize a Subgroup of Muscle-invasive Bladder Cancers with Excellent Response to Neoadjuvant Chemotherapy. *Eur Urol,* 69**,** 384-8.

GUICHARD, G., LARRE, S., GALLINA, A., LAZAR, A., FAUCON, H., CHEMAMA, S., ALLORY, Y., PATARD, J. J., VORDOS, D., HOZNEK, A., YIOU, R., SALOMON, L., ABBOU, C. C. & DE LA TAILLE, A. 2007. Extended 21-sample needle biopsy protocol for diagnosis of prostate cancer in 1000 consecutive patients. *Eur Urol,* 52**,** 430-5.

GUO, G., SUN, X., CHEN, C., WU, S., HUANG, P., LI, Z., DEAN, M., HUANG, Y., JIA, W., ZHOU, Q., TANG, A., YANG, Z., LI, X., SONG, P., ZHAO, X., YE, R., ZHANG, S., LIN, Z., QI, M., WAN, S., XIE, L., FAN, F., NICKERSON, M. L., ZOU, X., HU, X., XING, L., LV, Z., MEI, H., GAO, S., LIANG, C., GAO, Z., LU, J., YU, Y., LIU, C., LI, L., FANG, X., JIANG, Z., YANG, J., LI, C., ZHAO, X., CHEN, J., ZHANG, F., LAI, Y., LIN, Z., ZHOU, F., CHEN, H., CHAN, H. C., TSANG, S., THEODORESCU, D., LI, Y., ZHANG, X., WANG, J., YANG, H., GUI, Y., WANG, J. & CAI, Z. 2013. Whole-genome and whole-exome sequencing of bladder cancer identifies frequent alterations in genes involved in sister chromatid cohesion and segregation. *Nat Genet,* 45**,** 1459-63.

HA, T. T., HUY, N. T., MURAO, L. A., LAN, N. T., THUY, T. T., TUAN, H. M., NGA, C. T., TUONG, V. V., DAT, T. V., KIKUCHI, M., YASUNAMI, M., MORITA, K., HUONG, V. T. & HIRAYAMA, K. 2011. Elevated levels of cell-free circulating DNA in patients with acute dengue virus infection. *PLoS One,* 6**,** e25969.

HAAS, G. P., DELONGCHAMPS, N., BRAWLEY, O. W., WANG, C. Y. & DE LA ROZA, G. 2008. The worldwide epidemiology of prostate cancer: perspectives from autopsy studies. *Can J Urol,* 15**,** 3866-71.

HAFFNER, M. C., MOSBRUGER, T., ESOPI, D. M., FEDOR, H., HEAPHY, C. M., WALKER, D. A., ADEJOLA, N., GUREL, M., HICKS, J., MEEKER, A. K., HALUSHKA, M. K., SIMONS, J. W., ISAACS, W. B., DE MARZO, A. M., NELSON, W. G. & YEGNASUBRAMANIAN, S. 2013. Tracking the clonal origin of lethal prostate cancer. *J Clin Invest,* 123**,** 4918-22.

HAYDEN, E. C. 2014. Technology: The $1,000 genome. *Nature,* 507**,** 294-5.

HEIDENREICH, A., BASTIAN, P. J., BELLMUNT, J., BOLLA, M., JONIAU, S., VAN DER KWAST, T., MASON, M., MATVEEV, V., WIEGEL, T., ZATTONI, F., MOTTET, N. & EUROPEAN ASSOCIATION OF, U. 2014. EAU guidelines on prostate cancer. part 1: screening, diagnosis, and local treatment with curative intent-update 2013. *Eur Urol,* 65**,** 124-37.

HEITZER, E., ULZ, P., BELIC, J., GUTSCHI, S., QUEHENBERGER, F., FISCHEREDER, K., BENEZEDER, T., AUER, M., PISCHLER, C., MANNWEILER, S., PICHLER, M., EISNER, F., HAEUSLER, M., RIETHDORF, S., PANTEL, K., SAMONIGG, H., HOEFLER, G., AUGUSTIN, H., GEIGL, J. B. & SPEICHER, M. R. 2013. Tumor associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. *Genome Med,* 5**,** 30.

HESSELS, D., SMIT, F. P., VERHAEGH, G. W., WITJES, J. A., CORNEL, E. B. & SCHALKEN, J. A. 2007. Detection of TMPRSS2-ERG fusion transcripts and prostate cancer antigen 3 in urinary sediments may improve diagnosis of prostate cancer. *Clin Cancer Res,* 13**,** 5103-8.

HODGE, K. K., MCNEAL, J. E., TERRIS, M. K. & STAMEY, T. A. 1989. Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate. *J Urol,* 142**,** 71-4; discussion 74-5.

HODIS, E., WATSON, I. R., KRYUKOV, G. V., AROLD, S. T., IMIELINSKI, M., THEURILLAT, J. P., NICKERSON, E., AUCLAIR, D., LI, L., PLACE, C., DICARA, D., RAMOS, A. H., LAWRENCE, M. S., CIBULSKIS, K., SIVACHENKO, A., VOET, D., SAKSENA, G., STRANSKY, N., ONOFRIO, R. C., WINCKLER, W., ARDLIE, K., WAGLE, N., WARGO, J., CHONG, K., MORTON, D. L., STEMKE-HALE, K., CHEN, G., NOBLE, M., MEYERSON, M., LADBURY, J. E., DAVIES, M. A., GERSHENWALD, J. E., WAGNER, S. N., HOON, D. S., SCHADENDORF, D.,

LANDER, E. S., GABRIEL, S. B., GETZ, G., GARRAWAY, L. A. & CHIN, L. 2012. A landscape of driver mutations in melanoma. *Cell,* 150**,** 251-63.

HONG, M. K., MACINTYRE, G., WEDGE, D. C., VAN LOO, P., PATEL, K., LUNKE, S., ALEXANDROV, L. B., SLOGGETT, C., CMERO, M., MARASS, F., TSUI, D., MANGIOLA, S., LONIE, A., NAEEM, H., SAPRE, N., PHAL, P. M., KURGANOVS, N., CHIN, X., KERGER, M., WARREN, A. Y., NEAL, D., GNANAPRAGASAM, V., ROSENFELD, N., PEDERSEN, J. S., RYAN, A., HAVIV, I., COSTELLO, A. J., CORCORAN, N. M. & HOVENS, C. M. 2015. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. *Nat Commun,* 6**,** 6605.

HSING, A. W., TSAO, L. & DEVESA, S. S. 2000. International trends and patterns of prostate cancer incidence and mortality. *International Journal of Cancer,* 85**,** 60-67.

HURST, C. D., PLATT, F. M. & KNOWLES, M. A. 2014. Comprehensive mutation analysis of the TERT promoter in bladder cancer and detection of mutations in voided urine. *Eur Urol,* 65**,** 367-9.

INTERNATIONAL COLLABORATION OF, T., MEDICAL RESEARCH COUNCIL ADVANCED BLADDER CANCER WORKING, P., EUROPEAN ORGANISATION FOR, R., TREATMENT OF CANCER GENITO-URINARY TRACT CANCER, G., AUSTRALIAN BLADDER CANCER STUDY, G., NATIONAL CANCER INSTITUTE OF CANADA CLINICAL TRIALS, G., FINNBLADDER, NORWEGIAN BLADDER CANCER STUDY, G., CLUB UROLOGICO ESPANOL DE TRATAMIENTO ONCOLOGICO, G., GRIFFITHS, G., HALL, R., SYLVESTER, R., RAGHAVAN, D. & PARMAR, M. K. 2011. International phase III trial assessing neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: long-term results of the BA06 30894 trial. *J Clin Oncol,* 29**,** 2171-7.

JIANG, P., CHAN, C. W., CHAN, K. C., CHENG, S. H., WONG, J., WONG, V. W., WONG, G. L., CHAN, S. L., MOK, T. S., CHAN, H. L., LAI, P. B., CHIU, R. W. & LO, Y. M. 2015. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A,* 112**,** E1317-25.

JUNG, K., FLEISCHHACKER, M. & RABIEN, A. 2010. Cell-free DNA in the blood as a solid tumor biomarker—A critical appraisal of the literature. *Clinica Chimica Acta,* 411**,** 1611-1624.

KATTAN, M. W., EASTHAM, J. A., STAPLETON, A. M., WHEELER, T. M. & SCARDINO, P. T. 1998. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst,* 90**,** 766-71.

KENT, W. J., SUGNET, C. W., FUREY, T. S., ROSKIN, K. M., PRINGLE, T. H., ZAHLER, A. M. & HAUSSLER, D. 2002. The human genome browser at UCSC. *Genome Res,* 12**,** 996-1006.

KINDE, I., BETTEGOWDA, C., WANG, Y., WU, J., AGRAWAL, N., SHIH IE, M., KURMAN, R., DAO, F., LEVINE, D. A., GIUNTOLI, R., RODEN, R., ESHLEMAN, J. R., CARVALHO, J. P., MARIE, S. K., PAPADOPOULOS, N., KINZLER, K. W., VOGELSTEIN, B. & DIAZ, L. A., JR. 2013a. Evaluation of DNA from the Papanicolaou test to detect ovarian and endometrial cancers. *Sci Transl Med,* 5**,** 167ra4.

KINDE, I., MUNARI, E., FARAJ, S. F., HRUBAN, R. H., SCHOENBERG, M., BIVALACQUA, T., ALLAF, M., SPRINGER, S., WANG, Y., DIAZ, L. A., JR., KINZLER, K. W., VOGELSTEIN, B., PAPADOPOULOS, N. & NETTO, G. J. 2013b. TERT promoter mutations occur early in urothelial neoplasia and are biomarkers of early disease and disease recurrence in urine. *Cancer Res,* 73**,** 7162-7.

KINDE, I., WU, J., PAPADOPOULOS, N., KINZLER, K. W. & VOGELSTEIN, B. 2011. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A,* 108**,** 9530-5.

KING, C. R., MCNEAL, J. E., GILL, H. & PRESTI, J. C., JR. 2004. Extended prostate biopsy scheme improves reliability of Gleason grading: implications for radiotherapy patients. *Int J Radiat Oncol Biol Phys,* 59**,** 386-91.

KIRKALI, Z., CHAN, T., MANOHARAN, M., ALGABA, F., BUSCH, C., CHENG, L., KIEMENEY, L., KRIEGMAIR, M., MONTIRONI, R., MURPHY, W. M., SESTERHENN, I. A., TACHIBANA, M. & WEIDER, J. 2005. Bladder cancer: epidemiology, staging and grading, and diagnosis. *Urology,* 66**,** 4-34.

KLEIN, E. A., HADDAD, Z., YOUSEFI, K., LAM, L. L., WANG, Q., CHOEURNG, V., PALMER-ARONSTEN, B., BUERKI, C., DAVICIONI, E., LI, J., KATTAN, M. W., STEPHENSON, A. J. & MAGI-GALLUZZI, C. 2016. Decipher Genomic Classifier Measured on Prostate Biopsy Predicts Metastasis Risk. *Urology,* 90**,** 148-52.

KLOTZ, L., ZHANG, L., LAM, A., NAM, R., MAMEDOV, A. & LOBLAW, A. 2010. Clinical results of long-term follow-up of a large, active surveillance cohort with localized prostate cancer. *J Clin Oncol,* 28**,** 126-31.

KNEZEVIC, D., GODDARD, A. D., NATRAJ, N., CHERBAVAZ, D. B., CLARK-LANGONE, K. M., SNABLE, J., WATSON, D., FALZARANO, S. M., MAGI-GALLUZZI, C., KLEIN, E. A. & QUALE, C. 2013. Analytical

validation of the Oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies. *BMC Genomics,* 14**,** 690.

KOIE, T., OHYAMA, C., YAMAMOTO, H., IMAI, A., HATAKEYAMA, S., YONEYAMA, T., HASHIMOTO, Y., YONEYAMA, T. & TOBISAWA, Y. 2015. Differences in the recurrence pattern after neoadjuvant chemotherapy compared to surgery alone in patients with muscle-invasive bladder cancer. *Med Oncol,* 32**,** 421.

KUMAR, A., WHITE, T. A., MACKENZIE, A. P., CLEGG, N., LEE, C., DUMPIT, R. F., COLEMAN, I., NG, S. B., SALIPANTE, S. J., RIEDER, M. J., NICKERSON, D. A., COREY, E., LANGE, P. H., MORRISSEY, C., VESSELLA, R. L., NELSON, P. S. & SHENDURE, J. 2011. Exome sequencing identifies a spectrum of mutation frequencies in advanced and lethal prostate cancers. *Proceedings of the National Academy of Sciences,* 108**,** 17087-17092.

LAM, N. Y. L., RAINER, T. H., CHIU, R. W. K. & LO, Y. M. D. 2004. EDTA Is a Better Anticoagulant than Heparin or Citrate for Delayed Blood Processing for Plasma DNA Analysis. *Clinical Chemistry,* 50**,** 256.

LAMY, A., GOBET, F., LAURENT, M., BLANCHARD, F., VARIN, C., MOULIN, C., ANDREOU, A., FREBOURG, T. & PFISTER, C. 2006. Molecular profiling of bladder tumors based on the detection of FGFR3 and TP53 mutations. *J Urol,* 176**,** 2686-9.

LAUS, S., KINGSLEY, L. A., GREEN, M. & WADOWSKY, R. M. 2011. Comparison of QIAsymphony automated and QIAamp manual DNA extraction systems for measuring Epstein-Barr virus DNA load in whole blood using real-time PCR. *J Mol Diagn,* 13**,** 695-700.

LAXMAN, B., TOMLINS, S. A., MEHRA, R., MORRIS, D. S., WANG, L., HELGESON, B. E., SHAH, R. B., RUBIN, M. A., WEI, J. T. & CHINNAIYAN, A. M. 2006. Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer. *Neoplasia,* 8**,** 885-8.

LEARY, R. J., KINDE, I., DIEHL, F., SCHMIDT, K., CLOUSER, C., DUNCAN, C., ANTIPOVA, A., LEE, C., MCKERNAN, K., DE LA VEGA, F. M., KINZLER, K. W., VOGELSTEIN, B., DIAZ, L. A., JR. & VELCULESCU, V. E. 2010. Development of personalized tumor biomarkers using massively parallel sequencing. *Sci Transl Med,* 2**,** 20ra14.

LECORNET, E., AHMED, H. U., HU, Y., MOORE, C. M., NEVOUX, P., BARRATT, D., HAWKES, D., VILLERS, A. & EMBERTON, M. 2012. The accuracy of different biopsy strategies for the detection of clinically important prostate cancer: a computer simulation. *J Urol,* 188**,** 974-80.

LEE, F. C., HARRIS, W., CHENG, H. H., SHENOI, J., ZHAO, S., WANG, J., CHAMPION, T., IZARD, J., GORE, J. L., PORTER, M., YU, E. Y. & WRIGHT, J. L. 2013. Pathologic Response Rates of Gemcitabine/Cisplatin versus Methotrexate/Vinblastine/Adriamycin/Cisplatin Neoadjuvant Chemotherapy for Muscle Invasive Urothelial Bladder Cancer. *Advances in Urology,* 2013**,** 317190.

LEON, S. A., SHAPIRO, B., SKLAROFF, D. M. & YAROS, M. J. 1977. Free DNA in the serum of cancer patients and the effect of therapy. *Cancer Res,* 37**,** 646-50.

LI, H. 2011. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics,* 27**,** 718-9.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics,* 25**,** 2078-9.

LI, Y., ZHOU, X., ST JOHN, M. A. & WONG, D. T. 2004. RNA profiling of cell-free saliva using microarray technology. *J Dent Res,* 83**,** 199-203.

LILJA, H., ULMERT, D. & VICKERS, A. J. 2008. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. *Nat Rev Cancer,* 8**,** 268-78.

LIU, W., LAITINEN, S., KHAN, S., VIHINEN, M., KOWALSKI, J., YU, G., CHEN, L., EWING, C. M., EISENBERGER, M. A., CARDUCCI, M. A., NELSON, W. G., YEGNASUBRAMANIAN, S., LUO, J., WANG, Y., XU, J., ISAACS, W. B., VISAKORPI, T. & BOVA, G. S. 2009. Copy number analysis indicates monoclonal origin of lethal metastatic prostate cancer. *Nat Med,* 15**,** 559-65.

LO, Y. M., CHAN, K. C., SUN, H., CHEN, E. Z., JIANG, P., LUN, F. M., ZHENG, Y. W., LEUNG, T. Y., LAU, T. K., CANTOR, C. R. & CHIU, R. W. 2010. Maternal plasma DNA sequencing reveals the genome-wide genetic and mutational profile of the fetus. *Sci Transl Med,* 2**,** 61ra91.

LUI, Y. Y. & DENNIS, Y. M. 2002. Circulating DNA in plasma and serum: biology, preanalytical issues and diagnostic applications. *Clin Chem Lab Med,* 40**,** 962-8.

MANDEL, P. & METAIS, P. 1948. Les acides nucleiques du plasma sanguin chez l'homme. *C R Seances Soc Biol Fil,* 142**,** 241-3.

MANOHARAN, M., AYYATHURAI, R. & SOLOWAY, M. S. 2009. Radical cystectomy for urothelial carcinoma of the bladder: an analysis of perioperative and survival outcome. *BJU Int,* 104**,** 1227-32.

MARKOULATOS, P., SIAFAKAS, N. & MONCANY, M. 2002. Multiplex polymerase chain reaction: a practical approach. *J Clin Lab Anal,* 16**,** 47-51.

MARLEAU, A. M., CHEN, C. S., JOYCE, J. A. & TULLIS, R. H. 2012. Exosome removal as a therapeutic adjuvant in cancer. *J Transl Med,* 10**,** 134.

MCANINCH, J. W., LUE, T. F. & SMITH, D. R. 2013. *Smith and Tanagho's general urology / editors, Jack W. McAninch, Thomas F. Lue,* New York, McGraw-Hill Professional.

MCBRIDE, D. J., ORPANA, A. K., SOTIRIOU, C., JOENSUU, H., STEPHENS, P. J., MUDIE, L. J., HÄMÄLÄINEN, E., STEBBINGS, L. A., ANDERSSON, L. C., FLANAGAN, A. M., DURBECQ, V., IGNATIADIS, M., KALLIONIEMI, O., HECKMAN, C. A., ALITALO, K., EDGREN, H., FUTREAL, P. A., STRATTON, M. R. & CAMPBELL, P. J. 2010. Use of cancer specific genomic rearrangements to quantify disease burden in plasma from patients with solid tumors. *Genes, Chromosomes and Cancer,* 49**,** 1062-1069.

MCINERNEY, P., ADAMS, P. & HADI, M. Z. 2014. Error Rate Comparison during Polymerase Chain Reaction by DNA Polymerase. *Mol Biol Int,* 2014**,** 287430.

MILBURY, C. A., ZHONG, Q., LIN, J., WILLIAMS, M., OLSON, J., LINK, D. R. & HUTCHISON, B. 2014. Determining lower limits of detection of digital PCR assays for cancer-related gene mutations. *Biomol Detect Quantif,* 1**,** 8-22.

MILLER, D. C., GRUBER, S. B., HOLLENBECK, B. K., MONTIE, J. E. & WEI, J. T. 2006. Incidence of initial local therapy among men with lower-risk prostate cancer in the United States. *J Natl Cancer Inst,* 98**,** 1134-41.

MILLHOLLAND, J. M. L., S.; FERNANDEZ, C.A.; SHUBER, A.P. 2012. Detection of low frequency FGFR3 mutations in the urine of bladder cancer patients using next-generation deep sequencing. *Research and Reports in Urology* 4**,** 33-40.

MONITOR 2016. 2015/16 National Tariff Payment System. *In:* HEALTH., D. O. (ed.). London.

MONTIRONI, R., CHENG, L., LOPEZ-BELTRAN, A., SCARPELLI, M., MAZZUCCHELLI, R., MIKUZ, G., KIRKALI, Z. & MONTORSI, F. 2010. Original Gleason system versus 2005 ISUP modified Gleason system: the importance of indicating which system is used in the patient's pathology and clinical reports. *Eur Urol,* 58**,** 369-73.

MOORE, C. M., ROBERTSON, N. L., ARSANIOUS, N., MIDDLETON, T., VILLERS, A., KLOTZ, L., TANEJA, S. S. & EMBERTON, M. 2013. Image-guided prostate biopsy using magnetic resonance imaging-derived targets: a systematic review. *Eur Urol,* 63**,** 125-40.

MORRISON, C. D., LIU, P., WOLOSZYNSKA-READ, A., ZHANG, J., LUO, W., QIN, M., BSHARA, W., CONROY, J. M., SABATINI, L., VEDELL, P., XIONG, D., LIU, S., WANG, J., SHEN, H., LI, Y., OMILIAN, A. R., HILL, A., HEAD, K., GURU, K., KUNNEV, D., LEACH, R., ENG, K. H., DARLAK, C., HOEFLICH, C., VEERANKI, S., GLENN, S., YOU, M., PRUITT, S. C., JOHNSON, C. S. & TRUMP, D. L. 2014. Whole-genome sequencing identifies genomic heterogeneity at a nucleotide and chromosomal level in bladder cancer. *Proceedings of the National Academy of Sciences,* 111**,** E672-E681.

MOSTAFA, M. H., SHEWEITA, S. A. & O'CONNOR, P. J. 1999. Relationship between Schistosomiasis and Bladder Cancer. *Clinical Microbiology Reviews,* 12**,** 97-111.

MOULIERE, F. & ROSENFELD, N. 2015. Circulating tumor-derived DNA is shorter than somatic DNA in plasma. *Proceedings of the National Academy of Sciences,* 112**,** 3178-3179.

MURTAZA, M., DAWSON, S. J., TSUI, D. W., GALE, D., FORSHEW, T., PISKORZ, A. M., PARKINSON, C., CHIN, S. F., KINGSBURY, Z., WONG, A. S., MARASS, F., HUMPHRAY, S., HADFIELD, J., BENTLEY, D., CHIN, T. M., BRENTON, J. D., CALDAS, C. & ROSENFELD, N. 2013. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature,* 497**,** 108-12.

NEWMAN, A. M., BRATMAN, S. V., TO, J., WYNNE, J. F., ECLOV, N. C., MODLIN, L. A., LIU, C. L., NEAL, J. W., WAKELEE, H. A., MERRITT, R. E., SHRAGER, J. B., LOO, B. W., JR., ALIZADEH, A. A. & DIEHN, M. 2014. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat Med,* 20**,** 548-54.

NEWMAN, A. M., LOVEJOY, A. F., KLASS, D. M., KURTZ, D. M., CHABON, J. J., SCHERER, F., STEHR, H., LIU, C. L., BRATMAN, S. V., SAY, C., ZHOU, L., CARTER, J. N., WEST, R. B., SLEDGE, G. W., JR., SHRAGER, J. B., LOO, B. W., JR., NEAL, J. W., WAKELEE, H. A., DIEHN, M. & ALIZADEH, A. A. 2016. Integrated digital error suppression for improved detection of circulating tumor DNA. *Nat Biotechnol,* 34**,** 547-55.

NICE 2014. *Prostate Cancer: Diagnosis and Treatment.* Cardiff (UK).

NISHIMURA, K., FUJIYAMA, C., NAKASHIMA, K., SATOH, Y., TOKUDA, Y. & UOZUMI, J. 2009. The effects of neoadjuvant chemotherapy and chemo-radiation therapy on MRI staging in invasive bladder cancer: comparative study based on the pathological examination of whole layer bladder wall. *Int Urol Nephrol,* 41**,** 869-75.

NORTON, S. E., LUNA, K. K., LECHNER, J. M., QIN, J. & FERNANDO, M. R. 2013. A new blood collection device minimizes cellular DNA release during sample storage and shipping when compared to a standard device. *J Clin Lab Anal,* 27**,** 305-11.

ONS, Office For National Statistics. 2011. Cancer survival in England: Patients diagnosed 2005-2009 and followed up to 2010. *In:* ONS (ed.). London.

OTSUKA, J., OKUDA, T., SEKIZAWA, A., AMEMIYA, S., SAITO, H., OKAI, T. & KUSHIMA, M. 2004. Detection of p53 mutations in the plasma DNA of patients with ovarian cancer. *Int J Gynecol Cancer,* 14**,** 459-64.

PAN, W., GU, W., NAGPAL, S., GEPHART, M. H. & QUAKE, S. R. 2015. Brain tumor mutations detected in cerebral spinal fluid. *Clin Chem,* 61**,** 514-22.

PARPART-LI, S., BARTLETT, B., POPOLI, M., ADLEFF, V., TUCKER, L., STEINBERG, R., GEORGIADIS, A., PHALLEN, J., BRAHMER, J., AZAD, N., BROWNER, I., LAHERU, D., VELCULESCU, V. E., SAUSEN, M. & DIAZ, L. A., JR. 2017. The Effect of Preservative and Temperature on the Analysis of Circulating Tumor DNA. *Clin Cancer Res,* 23**,** 2471-2477.

PATEL, K. M. & TSUI, D. W. Y. 2015. The translational potential of circulating tumour DNA in oncology. *Clinical Biochemistry,* 48**,** 957-961.

PATEL, K. M., VAN DER VOS, K. E., SMITH, C. G., MOULIERE, F., TSUI, D., MORRIS, J., CHANDRANANDA, D., MARASS, F., VAN DEN BROEK, D., NEAL, D. E., GNANAPRAGASAM, V. J., FORSHEW, T., VAN RHIJN, B. W., MASSIE, C. E., ROSENFELD, N. & VAN DER HEIJDEN, M. S. 2017. Association Of Plasma And Urinary Mutant DNA With Clinical Outcomes In Muscle Invasive Bladder Cancer. *Sci Rep,* 7**,** 5554.

PENG, Q., VIJAYA SATYA, R., LEWIS, M., RANDAD, P. & WANG, Y. 2015. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genomics,* 16**,** 589.

PINTO, P. A., CHUNG, P. H., RASTINEHAD, A. R., BACCALA, A. A., JR., KRUECKER, J., BENJAMIN, C. J., XU, S., YAN, P., KADOURY, S., CHUA, C., LOCKLIN, J. K., TURKBEY, B., SHIH, J. H., GATES, S. P., BUCKNER, C., BRATSLAVSKY, G., LINEHAN, W. M., GLOSSOP, N. D., CHOYKE, P. L. & WOOD, B. J. 2011. Magnetic resonance imaging/ultrasound fusion guided prostate biopsy improves cancer detection following transrectal ultrasound biopsy and correlates with multiparametric magnetic resonance imaging. *J Urol,* 186**,** 1281-5.

POUND, C. R., PARTIN, A. W., EISENBERGER, M. A., CHAN, D. W., PEARSON, J. D. & WALSH, P. C. 1999. Natural history of progression after PSA elevation following radical prostatectomy. *JAMA,* 281**,** 1591-7.

PRATES, C., SOUSA, S., OLIVEIRA, C. & IKRAM, S. 2011. Prostate metastatic bone cancer in an Egyptian Ptolemaic mummy, a proposed radiological diagnosis. *International Journal of Paleopathology,* 1**,** 98-103.

QIAGEN. 2016. *Therascreen KRAS RGQ PCR Kit* [Online]. Available: https://www.qiagen.com/us/resources/technologies/oncology-companion-diagnostics/therascreen-kras-test-usa-labs/ - principle [Accessed 15/05/2017 2017].

QIAGEN. 2017. *QIAsymphony SP/AS instruments* [Online]. Available: https://www.qiagen.com/us/shop/automated-solutions/sample-preparation/qiasymphony-spas-instruments/ - productdetails [Accessed 21/03/2017].

QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13**,** 341.

R CORE TEAM 2015. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing

.

RAGO, C., HUSO, D. L., DIEHL, F., KARIM, B., LIU, G., PAPADOPOULOS, N., SAMUELS, Y., VELCULESCU, V. E., VOGELSTEIN, B., KINZLER, K. W. & DIAZ, L. A., JR. 2007. Serial assessment of human tumor burdens in mice by the analysis of circulating DNA. *Cancer Res,* 67**,** 9364-70.

RAJ, G. V. & BOCHNER, B. H. 2007. Radical Cystectomy and Lymphadenectomy for Invasive Bladder Cancer: Towards the Evolution of an Optimal Surgical Standard. *Seminars in Oncology,* 34**,** 110-121.

RAJINIKANTH, A., MANOHARAN, M., SOLOWAY, C. T., CIVANTOS, F. J. & SOLOWAY, M. S. 2008. Trends in Gleason score: concordance between biopsy and prostatectomy over 15 years. *Urology,* 72**,** 177-82.

REESE, A. C., PIERORAZIO, P. M., HAN, M. & PARTIN, A. W. 2012. Contemporary evaluation of the National Comprehensive Cancer Network prostate cancer risk classification system. *Urology,* 80**,** 1075-9.

ROBINSON, D., VAN ALLEN, E. M., WU, Y. M., SCHULTZ, N., LONIGRO, R. J., MOSQUERA, J. M., MONTGOMERY, B., TAPLIN, M. E., PRITCHARD, C. C., ATTARD, G., BELTRAN, H., ABIDA, W., BRADLEY, R. K., VINSON, J., CAO, X., VATS, P., KUNJU, L. P., HUSSAIN, M., FENG, F. Y., TOMLINS, S. A., COONEY, K. A., SMITH, D. C., BRENNAN, C., SIDDIQUI, J., MEHRA, R., CHEN, Y., RATHKOPF, D. E., MORRIS, M. J., SOLOMON, S. B., DURACK, J. C., REUTER, V. E., GOPALAN, A., GAO, J., LODA, M., LIS, R. T., BOWDEN, M., BALK, S. P., GAVIOLA, G., SOUGNEZ, C., GUPTA, M., YU, E. Y., MOSTAGHEL, E. A., CHENG, H. H., MULCAHY, H., TRUE, L. D., PLYMATE, S. R., DVINGE, H., FERRALDESCHI, R., FLOHR, P., MIRANDA, S., ZAFEIRIOU, Z., TUNARIU, N., MATEO, J., PEREZ-LOPEZ, R., DEMICHELIS, F., ROBINSON, B. D., SCHIFFMAN, M., NANUS, D. M., TAGAWA, S. T., SIGARAS, A., ENG, K. W., ELEMENTO, O., SBONER, A., HEATH, E. I., SCHER, H. I., PIENTA, K. J., KANTOFF, P., DE BONO, J. S., RUBIN, M. A., NELSON, P. S., GARRAWAY, L. A., SAWYERS, C. L. & CHINNAIYAN, A. M. 2015. Integrative clinical genomics of advanced prostate cancer. *Cell,* 161**,** 1215-28.

RODRIGUES, G., WARDE, P., PICKLES, T., CROOK, J., BRUNDAGE, M., SOUHAMI, L., LUKKA, H. & GENITOURINARY RADIATION ONCOLOGISTS OF, C. 2012. Pre-treatment risk stratification of prostate cancer patients: A critical review. *Can Urol Assoc J,* 6**,** 121-7.

ROSENBERG, J. E., HOFFMAN-CENSITS, J., POWLES, T., VAN DER HEIJDEN, M. S., BALAR, A. V., NECCHI, A., DAWSON, N., O'DONNELL, P. H., BALMANOUKIAN, A., LORIOT, Y., SRINIVAS, S., RETZ, M. M., GRIVAS, P., JOSEPH, R. W., GALSKY, M. D., FLEMING, M. T., PETRYLAK, D. P., PEREZ-GRACIA, J. L., BURRIS, H. A., CASTELLANO, D., CANIL, C., BELLMUNT, J., BAJORIN, D., NICKLES, D., BOURGON, R., FRAMPTON, G. M., CUI, N., MARIATHASAN, S., ABIDOYE, O., FINE, G. D. & DREICER, R. 2016. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet,* 387**,** 1909-20.

SAGE SCIENCE. 2017. *Pippin Prep* [Online]. Available: http://www.sagescience.com/products/pippin-prep/ [Accessed 01/04/2017].

SAKR, W. A., HAAS, G. P., CASSIN, B. F., PONTES, J. E. & CRISSMAN, J. D. 1993. The frequency of carcinoma and intraepithelial neoplasia of the prostate in young male patients. *J Urol,* 150**,** 379-85.

SCATTONI, V., ZLOTTA, A., MONTIRONI, R., SCHULMAN, C., RIGATTI, P. & MONTORSI, F. 2007. Extended and saturation prostatic biopsy in the diagnosis and characterisation of prostate cancer: a critical analysis of the literature. *Eur Urol,* 52**,** 1309-22.

SCHEININ, I., SIE, D., BENGTSSON, H., VAN DE WIEL, M. A., OLSHEN, A. B., VAN THUIJL, H. F., VAN ESSEN, H. F., EIJK, P. P., RUSTENBURG, F., MEIJER, G. A., REIJNEVELD, J. C., WESSELING, P., PINKEL, D., ALBERTSON, D. G. & YLSTRA, B. 2014. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Research,* 24**,** 2022-2032.

SCHRODER, F. H., HUGOSSON, J., ROOBOL, M. J., TAMMELA, T. L., CIATTO, S., NELEN, V., KWIATKOWSKI, M., LUJAN, M., LILJA, H., ZAPPA, M., DENIS, L. J., RECKER, F., BERENGUER, A., MAATTANEN, L., BANGMA, C. H., AUS, G., VILLERS, A., REBILLARD, X., VAN DER KWAST, T., BLIJENBERG, B. G., MOSS, S. M., DE KONING, H. J., AUVINEN, A. & INVESTIGATORS, E. 2009. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med,* 360**,** 1320-8.

SCHRODER, F. H. & ZAPPA, M. 2012. Prostate-specific antigen testing in Tyrol, Austria: prostate cancer mortality reduction was supported by an update with mortality data up to 2008. *Int J Public Health,* 57**,** 45-7.

SELVADURAI, E. D., SINGHERA, M., THOMAS, K., MOHAMMED, K., WOODE-AMISSAH, R., HORWICH, A., HUDDART, R. A., DEARNALEY, D. P. & PARKER, C. C. 2013. Medium-term outcomes of active surveillance for localised prostate cancer. *Eur Urol,* 64**,** 981-7.

SESHAN, V. & OLSHEN, A. 2016. DNAcopy: DNA copy number data analysis. R package verion 1.46.0 ed.

SEXTON, W. J., WIEGAND, L. R., CORREA, J. J., POLITIS, C., DICKINSON, S. I. & KANG, L. C. 2010. Bladder cancer: a review of non-muscle invasive disease. *Cancer Control,* 17**,** 256-68.

SHAH, A., RACHET, B., MITRY, E., COOPER, N., BROWN, C. M. & COLEMAN, M. P. 2008. Survival from bladder cancer in England and Wales up to 2001. *Br J Cancer,* 99 Suppl 1**,** S86-9.

SHARIAT, S. F., KATTAN, M. W., VICKERS, A. J., KARAKIEWICZ, P. I. & SCARDINO, P. T. 2009. CRITICAL REVIEW OF PROSTATE CANCER PREDICTIVE TOOLS. *Future oncology (London, England),* 5**,** 1555-1584.

SHEN, M. M. & ABATE-SHEN, C. 2010. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev,* 24**,** 1967-2000.

SHERWOOD, J. L., CORCORAN, C., BROWN, H., SHARPE, A. D., MUSILOVA, M. & KOHLMANN, A. 2016. Optimised Pre-Analytical Methods Improve KRAS Mutation Detection in Circulating Tumour DNA (ctDNA) from Patients with Non-Small Cell Lung Cancer (NSCLC). *PLoS One,* 11**,** e0150197.

SIDRANSKY, D., VON ESCHENBACH, A., TSAI, Y. C., JONES, P., SUMMERHAYES, I., MARSHALL, F., PAUL, M., GREEN, P., HAMILTON, S. R., FROST, P. & ET AL. 1991. Identification of p53 gene mutations in bladder cancers and urine samples. *Science,* 252**,** 706-9.

SIMON, M. A., LOKESHWAR, V. B. & SOLOWAY, M. S. 2003. Current bladder cancer tests: unnecessary or beneficial? *Critical Reviews in Oncology/Hematology,* 47**,** 91-107.

SJODAHL, G., LAUSS, M., GUDJONSSON, S., LIEDBERG, F., HALLDEN, C., CHEBIL, G., MANSSON, W., HOGLUND, M. & LINDGREN, D. 2011. A systematic study of gene mutations in urothelial carcinoma; inactivating mutations in TSC2 and PIK3R1. *PLoS One,* 6**,** e18583.

SOH, J., TOYOOKA, S., AOE, K., ASANO, H., ICHIHARA, S., KATAYAMA, H., HIRAKI, A., KIURA, K., AOE, M., SANO, Y., SUGI, K., SHIMIZU, N. & DATE, H. 2006. Usefulness of EGFR mutation screening in pleural fluid to predict the clinical outcome of gefitinib treated patients with lung cancer. *Int J Cancer,* 119**,** 2353-8.

SORENSON, G. D., PRIBISH, D. M., VALONE, F. H., MEMOLI, V. A., BZIK, D. J. & YAO, S. L. 1994. Soluble normal and mutated DNA sequences from single-copy genes in human blood. *Cancer Epidemiol Biomarkers Prev,* 3**,** 67-71.

SOZZI, G., ROZ, L., CONTE, D., MARIANI, L., ANDRIANI, F., VERDERIO, P. & PASTORINO, U. 2005. Effects of prolonged storage of whole plasma or isolated plasma DNA on the results of circulating DNA quantification assays. *J Natl Cancer Inst,* 97**,** 1848-50.

STEIN, J. & SKINNER, D. 2006. Radical cystectomy for invasive bladder cancer: long-term results of a standard procedure. *World Journal of Urology,* 24**,** 296-304.

STEINBERG, D. M., SAUVAGEOT, J., PIANTADOSI, S. & EPSTEIN, J. I. 1997. Correlation of prostate needle biopsy and radical prostatectomy Gleason grade in academic and community settings. *Am J Surg Pathol,* 21**,** 566-76.

STEVENSON, M. 2016. epiR: Tools for the Analysis of Epidemiological Data. R package version 0.9-79 ed.

STRECK. 2017. *Cell-Free DNA Urine Preserve* [Online]. Available: https://www.streck.com/product.aspx?p=Cell-Free DNA Urine Preserve [Accessed 01/03/2017].

SU, Y. H., WANG, M., BRENNER, D. E., NG, A., MELKONYAN, H., UMANSKY, S., SYNGAL, S. & BLOCK, T. M. 2004. Human Urine Contains Small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. *J Mol Diagn,* 6**,** 101-7.

SYLVESTER, R. J., VAN DER MEIJDEN, A., WITJES, J. A., JAKSE, G., NONOMURA, N., CHENG, C., TORRES, A., WATSON, R. & KURTH, K. H. 2005. High-grade Ta urothelial carcinoma and carcinoma in situ of the bladder. *Urology,* 66**,** 90-107.

TALY, V., PEKIN, D., EL ABED, A. & LAURENT-PUIG, P. 2012. Detecting biomarkers with microdroplet technology. *Trends Mol Med,* 18**,** 405-16.

THIERRY, A. R., MOULIERE, F., GONGORA, C., OLLIER, J., ROBERT, B., YCHOU, M., DEL RIO, M. & MOLINA, F. 2010. Origin and quantification of circulating DNA in mice with human colorectal cancer xenografts. *Nucleic Acids Res,* 38**,** 6159-75.

TOGNERI, F. S., WARD, D. G., FOSTER, J. M., DEVALL, A. J., WOJTOWICZ, P., ALYAS, S., VASQUES, F. R., OUMIE, A., JAMES, N. D., CHENG, K. K., ZEEGERS, M. P., DESHMUKH, N., O'SULLIVAN, B., TANIERE, P., SPINK, K. G., MCMULLAN, D. J., GRIFFITHS, M. & BRYAN, R. T. 2016. Genomic complexity of urothelial bladder cancer revealed in urinary cfDNA. *Eur J Hum Genet,* 24**,** 1167-74.

TOMLINS, S. A., RHODES, D. R., PERNER, S., DHANASEKARAN, S. M., MEHRA, R., SUN, X. W., VARAMBALLY, S., CAO, X., TCHINDA, J., KUEFER, R., LEE, C., MONTIE, J. E., SHAH, R. B., PIENTA, K. J., RUBIN, M. A. & CHINNAIYAN, A. M. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science,* 310**,** 644-8.

TRIALISTS, I. C. O. 1999. Neoadjuvant cisplatin, methotrexate, and vinblastine chemotherapy for muscle-invasive bladder cancer: a randomised controlled trial. International collaboration of trialists. *Lancet,* 354**,** 533-40.

TSUI, N. B., JIANG, P., CHOW, K. C., SU, X., LEUNG, T. Y., SUN, H., CHAN, K. C., CHIU, R. W. & LO, Y. M. 2012. High resolution size analysis of fetal DNA in the urine of pregnant women by paired-end massively parallel sequencing. *PLoS One,* 7**,** e48319.

TURGEON, G. A., SOUHAMI, L., CURY, F. L., FARIA, S. L., DUCLOS, M., STURGEON, J. & KASSOUF, W. 2014. Hypofractionated intensity modulated radiation therapy in combined modality treatment for bladder preservation in elderly patients with invasive bladder cancer. *Int J Radiat Oncol Biol Phys,* 88**,** 326-31.

UNTERGASSER, A., CUTCUTACHE, I., KORESSAAR, T., YE, J., FAIRCLOTH, B. C., REMM, M. & ROZEN, S. G. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res,* 40**,** e115.

VALE, C. 2003. Neoadjuvant chemotherapy in invasive bladder cancer: a systematic review and meta-analysis. *The Lancet,* 361**,** 1927-1934.

VAN ALLEN, E. M., WAGLE, N., STOJANOV, P., PERRIN, D. L., CIBULSKIS, K., MARLOW, S., JANE-VALBUENA, J., FRIEDRICH, D. C., KRYUKOV, G., CARTER, S. L., MCKENNA, A., SIVACHENKO, A., ROSENBERG, M., KIEZUN, A., VOET, D., LAWRENCE, M., LICHTENSTEIN, L. T., GENTRY, J. G., HUANG, F. W., FOSTEL, J., FARLOW, D., BARBIE, D., GANDHI, L., LANDER, E. S., GRAY, S. W., JOFFE, S., JANNE, P., GARBER, J., MACCONAILL, L., LINDEMAN, N., ROLLINS, B., KANTOFF, P., FISHER, S. A., GABRIEL, S., GETZ, G. & GARRAWAY, L. A. 2014. Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med,* 20**,** 682-8.

VAN DE WIEL, M. A., KIM, K. I., VOSSE, S. J., VAN WIERINGEN, W. N., WILTING, S. M. & YLSTRA, B. 2007. CGHcall: calling aberrations for array CGH tumor profiles. *Bioinformatics,* 23**,** 892-4.

VAN DER VAART, M. & PRETORIUS, P. J. 2010. Is the role of circulating DNA as a biomarker of cancer being prematurely overrated? *Clinical Biochemistry,* 43**,** 26-36.

VAN KESSEL, K. E. M., KOMPIER, L. C., DE BEKKER-GROB, E. W., ZUIVERLOON, T. C. M., VERGOUWE, Y., ZWARTHOFF, E. C. & STEYERBERG, E. W. 2012. FGFR3 mutation analysis on voided urine samples to reduce cystoscopies and cost in non-muscle invasive bladder cancer surveillance: a comparison of three different strategies. *The Journal of Urology,* 189**,** 1676-81.

VOGELSTEIN, B. & KINZLER, K. W. 1999. Digital PCR. *Proc Natl Acad Sci U S A,* 96**,** 9236-41.

VOGELSTEIN, B., PAPADOPOULOS, N., VELCULESCU, V. E., ZHOU, S., DIAZ, L. A., JR. & KINZLER, K. W. 2013. Cancer genome landscapes. *Science,* 339**,** 1546-58.

WANG, Y., SPRINGER, S., MULVEY, C. L., SILLIMAN, N., SCHAEFER, J., SAUSEN, M., JAMES, N., RETTIG, E. M., GUO, T., PICKERING, C. R., BISHOP, J. A., CHUNG, C. H., CALIFANO, J. A., EISELE, D. W., FAKHRY, C., GOURIN, C. G., HA, P. K., KANG, H., KIESS, A., KOCH, W. M., MYERS, J. N., QUON, H., RICHMON, J. D., SIDRANSKY, D., TUFANO, R. P., WESTRA, W. H., BETTEGOWDA, C., DIAZ, L. A., PAPADOPOULOS, N., KINZLER, K. W., VOGELSTEIN, B. & AGRAWAL, N. 2015. Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. *Science Translational Medicine,* 7**,** 293ra104-293ra104.

WARD, D. G., BAXTER, L., GORDON, N. S., OTT, S., SAVAGE, R. S., BEGGS, A. D., JAMES, J. D., LICKISS, J., GREEN, S., WALLIS, Y., WEI, W., JAMES, N. D., ZEEGERS, M. P., CHENG, K. K., MATHEWS, G. M., PATEL, P., GRIFFITHS, M. & BRYAN, R. T. 2016. Multiplex PCR and Next Generation Sequencing for the Non-Invasive Detection of Bladder Cancer. *PLoS One,* 11**,** e0149756.

WARREN, A. Y., WHITAKER, H. C., HAYNES, B., SANGAN, T., MCDUFFUS, L. A., KAY, J. D. & NEAL, D. E. 2013. Method for sampling tissue for research which preserves pathological data in radical prostatectomy. *Prostate,* 73**,** 194-202.

WEI, X., WALIA, V., LIN, J. C., TEER, J. K., PRICKETT, T. D., GARTNER, J., DAVIS, S., PROGRAM, N. C. S., STEMKE-HALE, K., DAVIES, M. A., GERSHENWALD, J. E., ROBINSON, W., ROBINSON, S., ROSENBERG, S. A. & SAMUELS, Y. 2011. Exome sequencing identifies GRIN2A as frequently mutated in melanoma. *Nat Genet,* 43**,** 442-6.

WILT, T. J., BRAWER, M. K., JONES, K. M., BARRY, M. J., ARONSON, W. J., FOX, S., GINGRICH, J. R., WEI, J. T., GILHOOLY, P., GROB, B. M., NSOULI, I., IYER, P., CARTAGENA, R., SNIDER, G., ROEHRBORN, C., SHARIFI, R., BLANK, W., PANDYA, P., ANDRIOLE, G. L., CULKIN, D., WHEELER, T. & PROSTATE CANCER INTERVENTION VERSUS OBSERVATION TRIAL STUDY, G. 2012. Radical prostatectomy versus observation for localized prostate cancer. *N Engl J Med,* 367**,** 203-13.

WITJES, J. A., COMPERAT, E., COWAN, N. C., DE SANTIS, M., GAKIS, G., LEBRET, T., RIBAL, M. J., VAN DER HEIJDEN, A. G., SHERIF, A. & EUROPEAN ASSOCIATION OF, U. 2014. EAU guidelines on muscle-invasive and metastatic bladder cancer: summary of the 2013 guidelines. *Eur Urol,* 65**,** 778-92.

WU, X. R. 2005. Urothelial tumorigenesis: a tale of divergent pathways. *Nat Rev Cancer,* 5**,** 713-25.

XUE, X., TEARE, M. D., HOLEN, I., ZHU, Y. M. & WOLL, P. J. 2009. Optimizing the yield and utility of circulating cell-free DNA from plasma and serum. *Clin Chim Acta,* 404**,** 100-4.

YAP, K. L., KIYOTANI, K., TAMURA, K., ANTIC, T., JANG, M., MONTOYA, M., CAMPANILE, A., YEW, P. Y., GANSHERT, C., FUJIOKA, T., STEINBERG, G. D., O'DONNELL, P. H. & NAKAMURA, Y. 2014. Whole-exome sequencing of muscle-invasive bladder cancer identifies recurrent mutations of UNC5C and prognostic importance of DNA repair gene mutations on survival. *Clin Cancer Res,* 20**,** 6605-17.

YEUNG, C., DINH, T. & LEE, J. 2014. The Health Economics of Bladder Cancer: An Updated Review of the Published Literature. *Pharmacoeconomics*.

YUNG, T. K., CHAN, K. C., MOK, T. S., TONG, J., TO, K. F. & LO, Y. M. 2009. Single-molecule detection of epidermal growth factor receptor mutations in plasma by microfluidics digital PCR in non-small cell lung cancer patients. *Clin Cancer Res,* 15**,** 2076-84.

ZEHNDER, P., STUDER, U. E., DANESHMAND, S., BIRKHAUSER, F. D., SKINNER, E. C., ROTH, B., MIRANDA, G., BURKHARD, F. C., CAI, J., SKINNER, D. G., THALMANN, G. N. & GILL, I. S. 2014. Outcomes of radical cystectomy with extended lymphadenectomy alone in patients with lymph node-positive bladder cancer who are unfit for or who decline adjuvant chemotherapy. *BJU Int,* 113**,** 554-60.

ZHANG, J., TONG, K. L., LI, P. K., CHAN, A. Y., YEUNG, C. K., PANG, C. C., WONG, T. Y., LEE, K. C. & LO, Y. M. 1999. Presence of donor- and recipient-derived DNA in cell-free urine samples of renal transplantation recipients: urinary DNA chimerism. *Clin Chem,* 45**,** 1741-6.

ZHANG, R., SHAO, F., WU, X. & YING, K. 2010. Value of quantitative analysis of circulating cell free DNA as a screening tool for lung cancer: a meta-analysis. *Lung Cancer,* 69**,** 225-31.

ZHANG, S., YU, Y. H., ZHANG, Y., QU, W. & LI, J. 2015. Radiotherapy in muscle-invasive bladder cancer: the latest research progress and clinical application. *Am J Cancer Res,* 5**,** 854-68.

ZUIVERLOON, T. C., VAN DER AA, M. N., VAN DER KWAST, T. H., STEYERBERG, E. W., LINGSMA, H. F., BANGMA, C. H. & ZWARTHOFF, E. C. 2010. Fibroblast growth factor receptor 3 mutation analysis on voided urine for surveillance of patients with low-grade non-muscle-invasive bladder cancer. *Clin Cancer Res,* 16**,** 3011-8.

# APPENDICES

## A-1 Amplifiable GE copies/ml of urine in cfDNA extracted from urine supernatant samples taken from healthy volunteers

| Sample ID | Volunteer ID | Sample Spun (Y/N) | DNA Extraction Kit | Extraction Sample Volume (ml) | Elution Volume (ul) | Enrichment Ratio | NA dPCR Loading Volume (ul) | Est. ROX Targets | Est. FAM Targets | Yield Est. | Amplifiable Copies/ml | Avearge Amplifiable copies /ml | Amplifiable copies / ul of eluate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNA | A | Y | Norgen | 4 | 50 | 1 | 1.5 | 182 | 168 | 0.554 | 3297.1 | 3890.4 | 311.2 |
| SNA | A | Y | Norgen | 4 | 50 | 1 | 1 | 165 | 165 | 0.587 | 4483.7 | | |
| SNB | B | Y | Norgen | 4 | 50 | 1 | 1.5 | 26 | 233 | 0.769 | 471.0 | 738.2 | 59.1 |
| SNB | B | Y | Norgen | 4 | 50 | 1 | 1 | 37 | 338 | 1.203 | 1005.4 | | |
| SNC | C | Y | Norgen | 4 | 50 | 1 | 1 | 31 | 242 | 0.861 | 842.4 | 656.7 | 52.5 |
| SNC | C | Y | Norgen | 4 | 50 | 1 | 1.5 | 26 | 268 | 0.884 | 471.0 | | |
| SNP | P | Y | Norgen | 1 | 50 | 1 | 1 | 40 | 261 | 0.929 | 4347.8 | 2934.8 | 58.7 |
| SNP | P | Y | Norgen | 1 | 50 | 1 | 1.5 | 21 | 170 | 0.561 | 1521.7 | | |
| SQA | A | Y | Qiagen | 4 | 50 | 1 | 1 | 200 | 296 | 1.053 | 5434.8 | 3890.4 | 311.2 |
| SQA | A | Y | Qiagen | 4 | 50 | 1 | 1.5 | 192 | 278 | 0.917 | 3478.3 | | |
| SQA | A | Y | Qiagen | 4 | 50 | 1 | 1.5 | 179 | 332 | 1.287 | 3242.8 | | |
| SQA | A | Y | Qiagen | 4 | 50 | 1 | 1.5 | 188 | 346 | 1.341 | 3405.8 | | |
| SQB | B | Y | Qiagen | 4 | 50 | 1 | 1.5 | 21 | 205 | 0.677 | 380.4 | 570.7 | 45.7 |
| SQB | B | Y | Qiagen | 4 | 50 | 1 | 1 | 28 | 240 | 0.854 | 760.9 | | |
| SQC | C | Y | Qiagen | 4 | 50 | 1 | 1 | 44 | 316 | 1.125 | 1195.7 | 933.0 | 74.6 |
| SQC | C | Y | Qiagen | 4 | 50 | 1 | 1.5 | 37 | 296 | 0.977 | 670.3 | | |
| SQP | P | Y | Qiagen | 1 | 50 | 1 | 1.5 | 102 | 299 | 1.191 | 7391.3 | 3768.1 | 75.4 |
| SQP | P | Y | Qiagen | 1 | 50 | 1 | 1.5 | 2 | 9 | 0.036 | 144.9 | | |
| SSnA | A | Y | SnoMag | 1 | 50 | 1 | 1.5 | 49 | 268 | 0.884 | 3550.7 | 3351.4 | 67.0 |
| SSnA | A | Y | SnoMag | 1 | 50 | 1 | 1 | 29 | 171 | 0.609 | 3152.2 | | |
| SSnB | B | Y | SnoMag | 1 | 50 | 1 | 1 | 38 | 219 | 0.779 | 4130.4 | 2391.3 | 47.8 |
| SSnB | B | Y | SnoMag | 1 | 50 | 1 | 1.5 | 9 | 243 | 0.802 | 652.2 | | |
| SSnC | C | Y | SnoMag | 1 | 50 | 1 | 1.5 | 4 | 157 | 0.518 | 289.9 | 634.1 | 12.7 |
| SSnC | C | Y | SnoMag | 1 | 50 | 1 | 1 | 9 | 192 | 0.683 | 978.3 | | |
| SSnP | P | Y | SnoMag | 1 | 50 | 1 | 1 | 9 | 201 | 0.715 | 978.3 | 1394.9 | 27.9 |
| SSnP | P | Y | SnoMag | 1 | 50 | 1 | 1.5 | 25 | 140 | 0.462 | 1811.6 | | |
| NB | B | N | Norgen | 4 | 50 | 1 | 1.5 | 296 | 0 | 0.000 | 5362.3 | 5615.9 | 449.3 |
| NB | B | N | Norgen | 4 | 50 | 1 | 1 | 307 | 0 | 0.000 | 5869.6 | | |
| NC | C | N | Norgen | 4 | 50 | 1 | 1.5 | 64 | 0 | 0.000 | 1159.4 | 1544.4 | 123.6 |
| NC | C | N | Norgen | 4 | 50 | 1 | 1 | 71 | 0 | 0.000 | 1929.3 | | |
| QB | B | N | Qiagen | 4 | 50 | 1 | 1.5 | 368 | 253 | 1.008 | 6666.7 | 8494.1 | 679.5 |
| QB | B | N | Qiagen | 4 | 50 | 1 | 1 | 429 | 297 | 1.669 | 11657.6 | | |
| QB | B | N | Qiagen | 4 | 50 | 1 | 1.5 | 463 | 319 | 1.236 | 8387.7 | | |
| QB | B | N | Qiagen | 4 | 50 | 1 | 1.5 | 401 | 280 | 1.085 | 7264.5 | | |
| QC | C | N | Qiagen | 4 | 50 | 1 | 1 | 103 | 270 | 1.076 | 1865.9 | 2237.3 | 179.0 |
| QC | C | N | Qiagen | 4 | 50 | 1 | 1.5 | 96 | 277 | 1.556 | 2608.7 | | |
| SnB | B | N | SnoMag | 1 | 50 | 1 | 1.5 | 49 | 215 | 0.857 | 3550.7 | 2155.8 | 43.1 |
| SnB | B | N | SnoMag | 4 | 50 | 1 | 1 | 28 | 139 | 0.781 | 760.9 | | |
| SnC | C | N | SnoMag | 1 | 50 | 1 | 1.5 | 10 | 80 | 0.319 | 724.6 | 403.1 | 8.1 |
| SnC | C | N | SnoMag | 4 | 50 | 1 | 1 | 3 | 44 | 0.247 | 81.5 | | |
| NE | E | N | Norgen | 4 | 50 | 1 | 1.5 | 316 | 151 | 0.619 | 5724.6 | 2343.0 | 187.4 |
| NE | E | N | Norgen | 4 | 50 | 1 | 1.5 | 375 | 217 | 0.882 | 6793.5 | | |
| NE | E | N | Norgen | 4 | 50 | 1 | 1.5 | 19 | 231 | 0.935 | 344.2 | | |
| NE | E | N | Norgen | 4 | 50 | 1 | 1.5 | 15 | 253 | 1.024 | 271.7 | | |
| NE | E | N | Norgen | 4 | 50 | 1 | 1.5 | 19 | 242 | 0.980 | 344.2 | | |
| NE | E | N | Norgen | 4 | 50 | 1 | 1.5 | 32 | 217 | 0.879 | 579.7 | | |
| SNE | E | Y | Norgen | 4 | 50 | 1 | 1.5 | 233 | 134 | 0.549 | 4221.0 | 5788.0 | 463.0 |
| SNE | E | Y | Norgen | 4 | 50 | 1 | 1.5 | 406 | 255 | 1.016 | 7355.1 | | |
| NF | F | N | Norgen | 4 | 50 | 1 | 1.5 | 445 | 163 | 0.668 | 8061.6 | 10221.9 | 817.8 |
| NF | F | N | Norgen | 4 | 50 | 1 | 1.5 | 675 | 244 | 0.992 | 12228.3 | | |
| NF | F | N | Norgen | 4 | 50 | 1 | 1.5 | 562 | 191 | 0.740 | 10181.2 | | |
| NF | F | N | Norgen | 4 | 50 | 1 | 1.5 | 575 | 199 | 0.771 | 10416.7 | | |
| NP | P | N | Norgen | 1 | 50 | 1 | 1.5 | 5 | 89 | 0.365 | 362.3 | 561.6 | 11.2 |
| NP | P | N | Norgen | 1 | 50 | 1 | 1.5 | 17 | 143 | 0.570 | 1231.9 | | |
| NP | P | N | Norgen | 1 | 50 | 1 | 1.5 | 6 | 0 | 0.000 | 434.8 | | |
| NP | P | N | Norgen | 1 | 50 | 1 | 1 | 2 | 0 | 0.000 | 217.4 | | |
| QE | E | N | Qiagen | 4 | 50 | 1 | 1.5 | 56 | 154 | 0.631 | 1014.5 | 878.6 | 70.3 |
| QE | E | N | Qiagen | 4 | 50 | 1 | 1.5 | 41 | 236 | 0.959 | 742.8 | | |
| QE | E | Y | Qiagen | 4 | 50 | 1 | 1.5 | 278 | 155 | 0.635 | 5036.2 | 8491.8 | 679.3 |
| QE | E | Y | Qiagen | 4 | 50 | 1 | 1.5 | 577 | 285 | 1.135 | 10452.9 | | |
| QE | E | Y | Qiagen | 4 | 50 | 1 | 1.5 | 522 | 267 | 1.035 | 9456.5 | | |
| QE | E | Y | Qiagen | 4 | 50 | 1 | 1.5 | 498 | 225 | 0.872 | 9021.7 | | |

# Appendices - Amplifiable GE copies/ml of urine in cfDNA extracted from urine supernatant samples taken from healthy volunteers

| Sample ID | Volunteer ID | Sample Spun (Y/N) | DNA Extraction Kit | Extraction Sample Volume (ml) | Elution Volume (ul) | Enrichment Ratio | NA dPCR Loading Volume (ul) | Est. ROX Targets | Est. FAM Targets | Yield Est. | Amplifiable Copies/ml | Avearge Amplifiable copies /ml | Amplifiable copies / ul of eluate |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| QF | F | N | Qiagen | 4 | 50 | 1 | 1.5 | 672 | 111 | 0.455 | 12173.9 | 13931.2 | 1114.5 |
| QF | F | N | Qiagen | 4 | 50 | 1 | 1.5 | 866 | 420 | 1.707 | 15688.4 | | |
| QP | P | N | Qiagen | 1 | 50 | 1 | 1.5 | 61 | 341 | 1.398 | 4420.3 | 2253.6 | 45.1 |
| QP | P | N | Qiagen | 1 | 50 | 1 | 1.5 | 57 | 251 | 1.020 | 1032.6 | | |
| QP | P | N | Qiagen | 1 | 50 | 1 | 1.5 | 31 | 300 | 1.195 | 2246.4 | | |
| QP | P | N | Qiagen | 1 | 50 | 1 | 1 | 26 | 287 | 1.612 | 2826.1 | | |
| QP | P | N | Qiagen | 1 | 50 | 1 | 1.5 | 41 | 124 | 0.620 | 742.8 | | |
| SnP | P | N | SnoMag | 1 | 50 | 1 | 1.5 | 7 | 87 | 0.347 | 507.2 | 416.7 | 8.3 |
| SnP | P | N | SnoMag | 1 | 50 | 1 | 1 | 3 | 88 | 0.494 | 326.1 | | |
| NeoE | E | N | NeoGeneStar | 5 | 50 | 1 | 1.5 | 200 | 138 | 0.566 | 2898.6 | 3333.3 | 333.3 |
| NeoE | E | N | NeoGeneStar | 5 | 50 | 1 | 1.5 | 260 | 150 | 0.610 | 3768.1 | | |
| SNeoE | E | Y | NeoGeneStar | 5 | 50 | 1 | 1.5 | 632 | 176 | 0.715 | 9159.4 | 8717.4 | 871.7 |
| SNeoE | E | Y | NeoGeneStar | 5 | 50 | 1 | 1.5 | 571 | 212 | 0.845 | 8275.4 | | |
| NeoF | F | N | NeoGeneStar | 5 | 50 | 1 | 1.5 | 824 | 104 | 0.426 | 11942.0 | 13731.9 | 1373.2 |
| NeoF | F | N | NeoGeneStar | 5 | 50 | 1 | 1.5 | 1071 | 203 | 0.825 | 15521.7 | | |
| NeoP | P | N | NeoGeneStar | 1 | 50 | 1 | 1.5 | 1 | 2 | 0.008 | 72.5 | 36.2 | 0.7 |
| NeoP | P | N | NeoGeneStar | 1 | 50 | 1 | 1.5 | 0 | 1 | 0.004 | 0.0 | | |
| SNF | F | Y | Norgen | 4 | 50 | 1 | 1.5 | 1059 | 185 | 0.737 | 19184.8 | 22056.2 | 1764.5 |
| SNF | F | Y | Norgen | 4 | 50 | 1 | 1.5 | 1376 | 235 | 0.955 | 24927.5 | | |
| SQF | F | Y | Qiagen | 4 | 50 | 4 | 1.5 | 586 | 73 | 1.000 | 10615.9 | 10534.4 | 842.8 |
| SQF | F | Y | Qiagen | 4 | 50 | 4 | 1.5 | 577 | 95 | 1.301 | 10452.9 | | |
| SNeoF | F | Y | NeoGeneStar | 5 | 50 | 4 | 1.5 | 748 | 38 | 0.521 | 10840.6 | 10398.6 | 1039.9 |
| SNeoF | F | Y | NeoGeneStar | 5 | 50 | 4 | 1.5 | 687 | 54 | 0.740 | 9956.5 | | |
| NG1 | G1 | N | Norgen | 4 | 50 | 1 | 1.5 | 976 | 211 | 0.977 | 17681.2 | 22537.6 | 1803.0 |
| NG1 | G1 | N | Norgen | 4 | 50 | 1 | 1.5 | 927 | 160 | 0.741 | 16793.5 | | |
| NG1 | G1 | N | Norgen | 4 | 50 | 1 | 1.5 | 878 | 160 | 0.741 | 15905.8 | | |
| NG1 | G1 | N | Norgen | 4 | 50 | 4 | 1.8 | 116 | | | 29743.6 | | |
| NG1 | G1 | N | Norgen | 4 | 50 | 4 | 1.8 | 127 | | | 32564.1 | | |
| QG1 | G1 | N | Qiagen | 4 | 50 | 4 | 1.5 | 976 | 25 | 0.32051282 | 70724.6 | 68731.9 | 5498.6 |
| QG1 | G1 | N | Qiagen | 4 | 50 | 4 | 1.5 | 921 | 36 | 0.46153846 | 66739.1 | | |
| NG2 | G2 | N | Norgen | 4 | 50 | 4 | 1.5 | 651 | 78 | 1 | 47173.9 | 37759.2 | 3020.7 |
| NG2 | G2 | N | Norgen | 4 | 50 | 4 | 1.5 | 577 | 52 | 0.66666667 | 41811.6 | | |
| NG2 | G2 | N | Norgen | 4 | 50 | 4 | 1.8 | 120 | | | 30769.2 | | |
| NG2 | G2 | N | Norgen | 4 | 50 | 4 | 1.8 | 122 | | | 31282.1 | | |
| NG3 | G3 | N | Norgen | 4 | 50 | 4 | 1.5 | 601 | 76 | 0.97435897 | 43550.7 | 41123.2 | 3289.9 |
| NG3 | G3 | N | Norgen | 4 | 50 | 4 | 1.5 | 534 | 60 | 0.76923077 | 38695.7 | | |
| NG4 | G4 | N | Norgen | 4 | 50 | 4 | 1.5 | 663 | 64 | 0.82051282 | 48043.5 | 49275.4 | 3942.0 |
| NG4 | G4 | N | Norgen | 4 | 50 | 4 | 1.5 | 697 | 61 | 0.78205128 | 50507.2 | | |
| QG2 | G2 | N | Qiagen | 4 | 50 | 4 | 1.5 | 904 | 41 | 0.52564103 | 65507.2 | 66956.5 | 5356.5 |
| QG2 | G2 | N | Qiagen | 4 | 50 | 4 | 1.5 | 944 | 42 | 0.53846154 | 68405.8 | | |
| QG3 | G3 | N | Qiagen | 4 | 50 | 4 | 1.5 | 921 | 43 | 0.55128205 | 66739.1 | 56396.3 | 4511.7 |
| QG3 | G3 | N | Qiagen | 4 | 50 | 4 | 1.5 | 851 | 23 | 0.29487179 | 61666.7 | | |
| QG3 | G3 | N | Qiagen | 4 | 50 | 4 | 1.8 | 188 | | | 48205.1 | | |
| QG3 | G3 | N | Qiagen | 4 | 50 | 4 | 1.8 | 191 | | | 48974.4 | | |
| QG3 | G4 | N | Qiagen | 4 | 50 | 4 | 1.5 | 1157 | 46 | 1 | 83840.6 | 87137.7 | 6971.0 |
| QG3 | G4 | N | Qiagen | 4 | 50 | 4 | 1.5 | 1248 | 45 | 0.97826087 | 90434.8 | | |
| SNG1 | G1 | Y | Norgen | 4 | 50 | 4 | 1.5 | 363 | 80 | 1.73913043 | 26304.3 | 26123.2 | 2089.9 |
| SNG1 | G1 | Y | Norgen | 4 | 50 | 4 | 1.5 | 358 | 71 | 1.54347826 | 25942.0 | | |
| SQG1 | G1 | Y | Qiagen | 4 | 50 | 4 | 1.5 | 614 | 39 | 0.84782609 | 44492.8 | 43550.7 | 3484.1 |
| SQG1 | G1 | Y | Qiagen | 4 | 50 | 4 | 1.5 | 588 | 37 | 0.80434783 | 42608.7 | | |

DNA was extracted from the samples using 4 commercially available kits and dPCR was conducted as previously described in section 2.4.2 to target *RPP30*. Average amplifiable copies were calculated from replicate samples. Initially, multiple replicates were tested before reducing the replicates to 2 per sample.

# A-2 Amplifiable GE copies per ml of urine in cfDNA extracted from urine supernatant samples taken from healthy volunteers

| Sample ID | Volunteer ID | Sample Processing Time (hrs) | Sample Spun (Y/N) | EDTA (Y/N) | Extraction Sample Volume (ml) | Elution Volume (ul) | Enrichment Ratio | NA dPCR Loading Volume (ul) | Est. ROX Targets | Est. FAM Targets | Yield Est. | Amplifiable Copies/ml | Avearge Amplifiable copies /ml |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| A1 | A | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 43 | 126 | 140.0% | 4352 | 3391 |
| A1 | A | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 24 | 158 | 175.6% | 2429 | |
| A10 | A | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 915 | 45 | 50.0% | 92611 | 91144 |
| A10 | A | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 886 | 49 | 54.4% | 89676 | |
| A11 | A | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 11 | 122 | 135.6% | 1113 | 1063 |
| A11 | A | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 10 | 132 | 146.7% | 1012 | |
| A2 | A | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 2195 | 56 | 62.2% | 222166 | 228846 |
| A2 | A | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 2327 | 51 | 56.7% | 235526 | |
| A3 | A | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 4 | 131 | 145.6% | 405 | 253 |
| A3 | A | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 1 | 136 | 151.1% | 101 | |
| A4 | A | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 2327 | 52 | 57.8% | 235526 | 235526 |
| A4 | A | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 2327 | 50 | 55.6% | 235526 | |
| A5 | A | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 0 | 167 | 185.6% | 0 | 51 |
| A5 | A | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 1 | 166 | 184.4% | 101 | |
| A6 | A | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 2327 | 45 | 50.0% | 235526 | 235526 |
| A6 | A | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 2327 | 59 | 65.6% | 235526 | |
| A7 | A | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 29 | 164 | 182.2% | 2935 | 2530 |
| A7 | A | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 21 | 192 | 213.3% | 2126 | |
| A8 | A | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 2327 | 59 | 65.6% | 235526 | 235526 |
| A8 | A | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 2327 | 69 | 76.7% | 235526 | |
| A9 | A | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 16 | 81 | 90.0% | 1619 | 1569 |
| A9 | A | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 15 | 81 | 90.0% | 1518 | |
| B1 | B | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 1 | 181 | 635.1% | 101 | 101 |
| B1 | B | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 1 | 160 | 561.4% | 101 | |
| B10 | B | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 4 | 38 | 118.8% | 405 | 304 |
| B10 | B | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 2 | 52 | 162.5% | 202 | |
| B11 | B | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 1 | 77 | 240.6% | 101 | 51 |
| B11 | B | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 72 | 225.0% | 0 | |
| B12 | B | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 0 | 50 | 156.3% | 0 | 101 |
| B12 | B | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 2 | 57 | 178.1% | 202 | |
| B2 | B | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 7 | 158 | 175.6% | 709 | 1164 |
| B2 | B | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 16 | 134 | 148.9% | 1619 | |
| B3 | B | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 1 | 151 | 167.8% | 101 | 152 |
| B3 | B | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 2 | 161 | 178.9% | 202 | |
| B4 | B | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 13 | 153 | 170.0% | 1316 | 962 |
| B4 | B | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 6 | 161 | 178.9% | 607 | |
| B5 | B | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 16 | 161 | 503.1% | 1619 | 1974 |
| B5 | B | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 23 | 162 | 506.3% | 2328 | |
| B6 | B | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 16 | 127 | 396.9% | 1619 | 1721 |
| B6 | B | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 18 | 127 | 396.9% | 1822 | |
| B7 | B | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 3 | 127 | 396.9% | 304 | 253 |
| B7 | B | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 2 | 155 | 484.4% | 202 | |
| B8 | B | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 31 | 169 | 528.1% | 3138 | 3087 |
| B8 | B | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 30 | 123 | 384.4% | 3036 | |
| B9 | B | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 0 | 54 | 168.8% | 0 | 51 |
| B9 | B | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 1 | 79 | 246.9% | 101 | |
| C1 | C | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 3 | 164 | 575.4% | 304 | 304 |
| C1 | C | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 3 | 168 | 589.5% | 304 | |
| C10 | C | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 1 | 100 | 312.5% | 101 | 202 |
| C10 | C | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 3 | 92 | 287.5% | 304 | |
| C11 | C | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 119 | 371.9% | 0 | 152 |
| C11 | C | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 3 | 128 | 400.0% | 304 | |
| C12 | C | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 3 | 108 | 189.5% | 304 | 202 |
| C12 | C | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 1 | 85 | 149.1% | 101 | |
| C2 | C | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 14 | 122 | 214.0% | 1417 | 1468 |
| C2 | C | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 15 | 147 | 257.9% | 1518 | |
| C3 | C | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 20 | 150 | 468.8% | 2024 | 2277 |
| C3 | C | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 25 | 128 | 400.0% | 2530 | |
| C4 | C | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 11 | 138 | 431.3% | 1113 | 962 |
| C4 | C | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 8 | 140 | 437.5% | 810 | |
| C5 | C | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 2 | 140 | 437.5% | 202 | 253 |
| C5 | C | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 3 | 126 | 393.8% | 304 | |
| C6 | C | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 6 | 150 | 468.8% | 607 | 1215 |
| C6 | C | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 18 | 139 | 434.4% | 1822 | |
| C7 | C | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 2 | 129 | 403.1% | 202 | 304 |
| C7 | C | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 4 | 136 | 425.0% | 405 | |
| C8 | C | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 2 | 136 | 425.0% | 202 | 354 |
| C8 | C | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 5 | 131 | 409.4% | 506 | |
| C9 | C | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 3 | 129 | 403.1% | 304 | 152 |
| C9 | C | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 0 | 162 | 506.3% | 0 | |

# Appendices - Amplifiable GE copies per ml of urine in cfDNA extracted from urine supernatant samples taken from healthy volunteers

| Sample ID | Volunteer ID | Sample Processing Time (hrs) | Sample Spun (Y/N) | EDTA (Y/N) | Extraction Sample Volume (ml) | Elution Volume (ul) | Enrichment Ratio | NA dPCR Loading Volume (ul) | Est. ROX Targets | Est. FAM Targets | Yield Est. | Amplifiable Copies/ml | Avearge Amplifiable copies /ml |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D1 | D | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 8 | 155 | 543.9% | 810 | 810 |
| D1 | D | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 8 | 142 | 498.2% | 810 | |
| D10 | D | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 44 | 169 | 187.8% | 4453 | 3998 |
| D10 | D | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 35 | 139 | 154.4% | 3543 | |
| D11 | D | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 162 | 180.0% | 0 | 0 |
| D11 | D | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 139 | 154.4% | 0 | |
| D12 | D | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 4 | 68 | 75.6% | 405 | 354 |
| D12 | D | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 3 | 69 | 76.7% | 304 | |
| D2 | D | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 44 | 201 | 705.3% | 4453 | 5010 |
| D2 | D | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 55 | 155 | 543.9% | 5567 | |
| D3 | D | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 6 | 187 | 656.1% | 607 | 405 |
| D3 | D | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 2 | 158 | 554.4% | 202 | |
| D4 | D | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 49 | 185 | 649.1% | 4960 | 5162 |
| D4 | D | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 53 | 145 | 508.8% | 5364 | |
| D5 | D | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 5 | 172 | 603.5% | 506 | 506 |
| D5 | D | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 5 | 157 | 550.9% | 506 | |
| D6 | D | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 42 | 181 | 635.1% | 4251 | 4251 |
| D6 | D | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 42 | 151 | 529.8% | 4251 | |
| D7 | D | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 162 | 568.4% | 0 | 51 |
| D7 | D | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 1 | 145 | 508.8% | 101 | |
| D8 | D | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 30 | 138 | 431.3% | 3036 | 3138 |
| D8 | D | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 32 | 113 | 353.1% | 3239 | |
| D9 | D | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 8 | 149 | 465.6% | 810 | 709 |
| D9 | D | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 6 | 169 | 528.1% | 607 | |
| E1 | E | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 2 | 132 | 463.2% | 202 | 101 |
| E1 | E | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 0 | 147 | 515.8% | 0 | |
| E10 | E | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 65 | 201 | 223.3% | 6579 | 6123 |
| E10 | E | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 56 | 209 | 232.2% | 5668 | |
| E11 | E | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 162 | 180.0% | 0 | 0 |
| E11 | E | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 181 | 201.1% | 0 | |
| E12 | E | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 14 | 147 | 163.3% | 1417 | 1215 |
| E12 | E | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 10 | 151 | 167.8% | 1012 | |
| E2 | E | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 174 | 194 | 680.7% | 17611 | 18674 |
| E2 | E | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 195 | 151 | 529.8% | 19737 | |
| E3 | E | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 162 | 568.4% | 0 | 0 |
| E3 | E | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 172 | 603.5% | 0 | |
| E4 | E | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 33 | 169 | 593.0% | 3340 | 2935 |
| E4 | E | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 25 | 183 | 642.1% | 2530 | |
| E5 | E | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 3 | 178 | 624.6% | 304 | 202 |
| E5 | E | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 1 | 158 | 554.4% | 101 | |
| E6 | E | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 57 | 145 | 508.8% | 5769 | 5769 |
| E6 | E | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 57 | 185 | 649.1% | 5769 | |
| E7 | E | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 172 | 603.5% | 0 | 0 |
| E7 | E | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 0 | 174 | 610.5% | 0 | |
| E8 | E | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 37 | 136 | 238.6% | 3745 | 3543 |
| E8 | E | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 33 | 125 | 219.3% | 3340 | |
| E9 | E | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 0 | 135 | 421.9% | 0 | 0 |
| E9 | E | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 0 | 136 | 425.0% | 0 | |
| F1 | F | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 65 | 160 | 177.8% | 6579 | 7186 |
| F1 | F | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 77 | 140 | 155.6% | 7794 | |
| F10 | F | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 2 | 2 | 2.2% | 202 | 202 |
| F10 | F | 48 | N | Y | 3.8 | 75 | 1 | 1.8 | 2 | 2 | 2.2% | 202 | |
| F11 | F | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 2 | 136 | 151.1% | 202 | 152 |
| F11 | F | 48 | Y | N | 3.8 | 75 | 1 | 1.8 | 1 | 133 | 147.8% | 101 | |
| F12 | F | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 26 | 127 | 445.6% | 2632 | 2530 |
| F12 | F | 48 | N | N | 3.8 | 75 | 1 | 1.8 | 24 | 158 | 554.4% | 2429 | |
| F2 | F | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 220 | 160 | 561.4% | 22267 | 23836 |
| F2 | F | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 251 | 151 | 529.8% | 25405 | |
| F3 | F | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 18 | 197 | 691.2% | 1822 | 1721 |
| F3 | F | 1 | Y | N | 3.8 | 75 | 1 | 1.8 | 16 | 177 | 621.1% | 1619 | |
| F4 | F | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 167 | 169 | 593.0% | 16903 | 17358 |
| F4 | F | 1 | N | N | 3.8 | 75 | 1 | 1.8 | 176 | 176 | 617.5% | 17814 | |
| F5 | F | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 76 | 181 | 635.1% | 7692 | 8350 |
| F5 | F | 6 | Y | Y | 3.8 | 75 | 1 | 1.8 | 89 | 176 | 617.5% | 9008 | |
| F6 | F | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 251 | 167 | 586.0% | 25405 | 24545 |
| F6 | F | 6 | N | Y | 3.8 | 75 | 1 | 1.8 | 234 | 151 | 529.8% | 23684 | |
| F7 | F | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 21 | 178 | 624.6% | 2126 | 1619 |
| F7 | F | 6 | Y | N | 3.8 | 75 | 1 | 1.8 | 11 | 181 | 635.1% | 1113 | |
| F8 | F | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 69 | 142 | 443.8% | 6984 | 6883 |
| F8 | F | 6 | N | N | 3.8 | 75 | 1 | 1.8 | 67 | 153 | 478.1% | 6781 | |
| F9 | F | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 61 | 123 | 384.4% | 6174 | 6630 |
| F9 | F | 48 | Y | Y | 3.8 | 75 | 1 | 1.8 | 70 | 120 | 375.0% | 7085 | |
| G1 | G | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 303 | 120 | 210.5% | 30668 | 32540 |
| G1 | G | 1 | Y | Y | 3.8 | 75 | 1 | 1.8 | 340 | 103 | 180.7% | 34413 | |
| G2 | G | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 158 | 135 | 236.8% | 15992 | 16447 |
| G2 | G | 1 | N | Y | 3.8 | 75 | 1 | 1.8 | 167 | 126 | 221.1% | 16903 | |

DNA was extracted from the samples using Qiagen Circulating Nucleic acid commercially available kit. dPCR was conducted as previously described in section 2.4.2 to target *RPP30*. Average amplifiable copies were calculated from replicate samples. Of note, the yield is particularly high, however, this is likely to be due to reduced amount of *Xenopus Tropicalis* DNA being added to the 8 no-extraction controls that were used as comparisons. The median estimated targets detected for the no-extraction samples was only 32 (range 21-59) during this analysis or 164 GE copies / µL of eluate.

## A-3 Pilot Bladder Primer Panel

| Primer Name | Forward Sequence | Reverse Sequence | chr | Amplicon Start | Amplicon ENd |
|---|---|---|---|---|---|
| TP53_D0008_001 | GTGGGGAACAAGAAGTGGAGA | CTCCCTGCTTCTGTCTCCTAC | 17 | 7572903 | 7573031 |
| TP53_D0008_002 | AGGAAGGGGCTGAGGTCA | GCGCTTCGAGATGTTCCGA | 17 | 7573904 | 7574019 |
| TP53_D0008_003 | TTGAGTTCCAAGGCCTCATTCA | GTACTGTGTATATACTTACTTCTCCCCC | 17 | 7573975 | 7574077 |
| TP53_D0008_004 | AAACGGCATTTTGAGTGTTAGAC | ACAACACCAGCTCCTCTCC | 17 | 7576789 | 7576917 |
| TP53_D0008_005 | CATCCAGTGGTTTCTTCTTTGG | TTCACTTTTATCACCTTTCCTTGCC | 17 | 7576873 | 7576961 |
| TP53_D0008_006 | CGCTTCTTGTCCTGCTTGC | TGTGCCTGTCCTGGGAGA | 17 | 7576996 | 7577115 |
| TP53_D0008_007 | ATTCTCTTCCTCTGTGCGCC | TTGCTTCTCTTTTCCTATCCTGAGT | 17 | 7577074 | 7577182 |
| TP53_D0008_008 | GGTCAGAGGCAAGCAGAGG | CCTCACCATCATCACACTGGAA | 17 | 7577434 | 7577528 |
| TP53_D0008_009 | CAAGTGGCTCCTGACCTGG | CTAGGTTGGCTCTGACTGTACC | 17 | 7577484 | 7577612 |
| TP53_D0008_010 | ACTGTTACACATGTAGTTGTAGTGGA | CTTGCCACAGGTCTCCCC | 17 | 7577561 | 7577667 |
| TP53_D0008_011 | GGCCACTGACAACCACCC | TCGACATAGTGTGGTGGTGC | 17 | 7578120 | 7578213 |
| TP53_D0008_012 | AGACCTCAGGCGGCTCA | CCTCAGCATCTTATCCGAGTGG | 17 | 7578173 | 7578278 |
| TP53_D0008_013 | ATCCAAATACTCCACACGCAAATTT | CCTCTGATTCCTCACTGATTGCT | 17 | 7578228 | 7578317 |
| TP53_D0008_014 | TGTCGTCTCTCCAGCCCC | GGCCATCTACAAGCAGTCACA | 17 | 7578343 | 7578450 |
| TP53_D0008_015 | GCCTCACAACCTCCGTCAT | AACTGGCCAAGACCTGCC | 17 | 7578407 | 7578523 |
| TP53_D0008_016 | GGAATCAACCCACAGCTGC | AACTCTGTCTCCTTCCTCTTCCTA | 17 | 7578483 | 7578581 |
| TP53_D0008_017 | TGGAAGCCAGCCCCTCA | GTCATCTTCTGTCCCTTCCCAG | 17 | 7579284 | 7579408 |
| TP53_D0008_018 | CCGTAGCTGCCCTGGTAG | AGCTCCTACACCGGCG | 17 | 7579364 | 7579453 |
| TP53_D0008_019 | TGACAGGGGCCAGGAG | CAATGGTTCACTGAAGACCCA | 17 | 7579405 | 7579533 |
| TP53_D0008_020 | GCATTCTGGGAGCTTCATCTG | TGACTGCTCTTTTCACCCATCT | 17 | 7579488 | 7579616 |
| TP53_D0008_021 | CCCAGCCCAACCCTTGTC | GAAGCGAAAATTCCATGGGACTG | 17 | 7579677 | 7579766 |
| TP53_D0008_022 | GCCCTTCCAATGGATCCACT | CTTCCGGGTCACTGCCAT | 17 | 7579816 | 7579928 |
| TP53_D0008_023 | CTAGGATCTGACTGCGGCTC | TGGAAGTGTCTCATGCTGGATC | 17 | 7579887 | 7579981 |
| BRAF_D0008_001 | ACTGTTCAAACTGATGGGACCC | ACTTACTACACCTCAGATATATTTCTTCATGA | 7 | 140453098 | 140453209 |
| BRAF_D0009_001 | TCACCTATTTTTACTGTGAGGTCTTCA | TCTTCATAATGCTTGCTCTGATAGGA | 7 | 140453154 | 140453259 |

| | | | | | |
|---|---|---|---|---|---|
| BRAF_D0009_002 | CACATTACATACTTACCATGCCACTTT | GGACTCGAGTGATGATTGGGAG | 7 | 140481360 | 140481476 |
| PIK3CA_D0012_002 | CTCAAGAAGCAGAAAGGGAAGAATTT | TCTTTTCTTCACGGTTGCCTACT | 3 | 178916834 | 178916947 |
| PIK3CA_D0012_001 | CTTCGGCTTTTTCAACCCTTTTTAAAA | AATATTTTAGAAAGGGACAACAGTTAAGCT | 3 | 178916887 | 178917036 |
| PIK3CA_D0012_003 | CATTTCCACAGCTACACCATATATGAAT | ATTTGACTTTACCTTATCAATGTCTCGAAT | 3 | 178921448 | 178921589 |
| PIK3CA_D0012_004 | CCCATTATTATAGAGATGATTGTTGAATTTTCC | AACAAGTTTATATTTCCCCATGCCAAT | 3 | 178927868 | 178928011 |
| PIK3CA_D0008_001 | AGGGAAAATGACAAAGAACAGC | TTTTAGCACTTACCTGTGacTCCA | 3 | 178936028 | 178936135 |
| PIK3CA_D0012_005 | GATGCTTGGCTCTGGAATGC | TTGTCCAGCCACCATGATGT | 3 | 178951957 | 178952102 |
| PIK3CA_D0008_002 | CAAGAGGCTTTGGAGTATTTCATGAA | ATGCATGCTGTTTAATTGTGTGGA | 3 | 178952042 | 178952144 |
| KRAS_D0034_001 | CAGTCCTCATGTACTGGTCCC | TACAGGAAGCAAGTAGTAATTGATGGA | 12 | 25380235 | 25380340 |
| KRAS_D0034_002 | GATTCTGAATTAGCTGTATCGTCAAGG | GGCCTGCTGAAAATGACTGAATATAA | 12 | 25398240 | 25398330 |
| KRAS_D0036_002 | CTTGCTAAGTCCTGAGCCTGTT | TTAAGGACTCTGAAGATGTACCTATGG | 12 | 25378594 | 25378690 |
| KRAS_D0036_003 | TCTGTATTTATTTCAGTGTTACTTACCTGTC | TGATTTGCCTTCTAGAACAGTAGACA | 12 | 25378522 | 25378644 |
| FGFR_EX_6 | GCCCCTGAGCGTCATCT | TGCGTCACTGTACACCTTGC | 4 | 1803535 | 1803668 |
| FGFR_EX_8 | GAGCTGGTGGAGGCTGAC | GAGCCCAGGCCTTTCTTG | 4 | 1806067 | 1806207 |
| FGFR_EX_13 | CGAGGACAACGTGATGAAGA | GGCGTCCTACTGGCATGA | 4 | 1807821 | 1807953 |
| NFE2L2_D0048_003 | TTCTGACTGGATGTGCTGGG | CAAAAGGAGCAAGAGAAAGCCTTT | 2 | 178098760 | 178098858 |
| NFE2L2_D0048_001 | GTCAAATACTTCTCGACTTACTCCAA | GCATAATGTGAATTAATTTATGTGGTATCTGTC | 2 | 178098931 | 178099080 |
| NFE2L2_D0048_002 | GGAGTTGTTCTTGTCTTTCCTTTTCA | GATTGACATACTTTGGAGGCAAGATATAG | 2 | 178098858 | 178098988 |
| CTNNB1_D0008_001 | AGCGGCTGTTAGTCACTGG | GTATCCACATCCTCTTCCTCAGG | 3 | 41266060 | 41266179 |
| CTNNB1_SUPP | CCACTACCACAGCTCCTTCT | CTCAAAACTGCATTCTGACTTTCA | 3 | 41266119 | 41266296 |
| HRAS_D0036_001 | CCCCGGTGCGCATGTA | AGCAGGTGGTCATTGATGGG | 11 | 533830 | 533931 |
| HRAS_D0043_003 | TGCTGGCACCTGGACG | TAAGCTGGTGGTGGTGGGCG | 11 | 534180 | 534311 |

# A-4 Amplifiable GE copies in cfDNA extracted from urine supernatant samples taken from patients with metastatic bladder cancer

| Sample ID | Extraction Sample Volume (ml) | elution Volume (ul) | Enrichment Ratio | NA dPCR Loading Volume (ul) | Est. ROX Targets | Est. FAM Targets | Yield Est. | Amplifiable Copies/ml | Avearge Amplifiable copies /ml | Amplifiable copies / ul of eluate |
|---|---|---|---|---|---|---|---|---|---|---|
| K1 | 3.8 | 75 | 1 | 1.8 | 467 | 104 | 82.9% | 47267 | 47166 | 2390 |
| K1 | 3.8 | 75 | 1 | 1.8 | 465 | 95 | 75.7% | 47065 | | |
| K2 | 3.8 | 75 | 1 | 1.8 | 1701 | 37 | 29.5% | 172166 | 168219 | 8523 |
| K2 | 3.8 | 75 | 1 | 1.8 | 1623 | 39 | 31.1% | 164271 | | |
| K3 | 3.8 | 75 | 1 | 1.8 | 305 | 116 | 92.4% | 30870 | 31883 | 1615 |
| K3 | 3.8 | 75 | 1 | 1.8 | 325 | 95 | 75.7% | 32895 | | |
| K4 | 3.8 | 75 | 1 | 1.8 | 2327 | 46 | 36.7% | 235526 | 231579 | 11733 |
| K4 | 3.8 | 75 | 1 | 1.8 | 2249 | 56 | 44.6% | 227632 | | |
| K5 | 3.8 | 75 | 1 | 1.8 | 617 | 77 | 61.4% | 62449 | 60577 | 3069 |
| K5 | 3.8 | 75 | 1 | 1.8 | 580 | 83 | 66.1% | 58704 | | |
| K6 | 3.8 | 75 | 1 | 1.8 | 1320 | 47 | 37.5% | 133603 | 136842 | 6933 |
| K6 | 3.8 | 75 | 1 | 1.8 | 1384 | 52 | 41.4% | 140081 | | |
| K7 | 3.8 | 75 | 1 | 1.8 | 339 | 112 | 89.2% | 34312 | 33907 | 1718 |
| K7 | 3.8 | 75 | 1 | 1.8 | 331 | 114 | 90.8% | 33502 | | |
| K8 | 3.8 | 75 | 1 | 1.8 | 1808 | 46 | 36.7% | 182996 | 180972 | 9169 |
| K8 | 3.8 | 75 | 1 | 1.8 | 1768 | 59 | 47.0% | 178947 | | |
| K9 | 3.8 | 75 | 1 | 1.8 | 1036 | 75 | 59.8% | 104858 | 103138 | 5226 |
| K9 | 3.8 | 75 | 1 | 1.8 | 1002 | 77 | 61.4% | 101417 | | |
| K10 | 3.8 | 75 | 1 | 1.8 | 831 | 40 | 31.9% | 84109 | 81123 | 4110 |
| K10 | 3.8 | 75 | 1 | 1.8 | 772 | 45 | 35.9% | 78138 | | |
| K11 | 3.8 | 75 | 1 | 1.8 | 106 | 78 | 62.2% | 10729 | 11387 | 577 |
| K11 | 3.8 | 75 | 1 | 1.8 | 119 | 82 | 65.3% | 12045 | | |
| K12 | 3.8 | 75 | 1 | 1.8 | 1188 | 29 | 23.1% | 120243 | 121457 | 6154 |
| K12 | 3.8 | 75 | 1 | 1.8 | 1212 | 41 | 32.7% | 122672 | | |
| L1 | 3.8 | 75 | 1 | 1.8 | 111 | 132 | 105.2% | 11235 | 11032 | 559 |
| L1 | 3.8 | 75 | 1 | 1.8 | 107 | 134 | 106.8% | 10830 | | |
| L2 | 3.8 | 75 | 1 | 1.8 | 2327 | 52 | 41.4% | 235526 | 235526 | 11933 |
| L2 | 3.8 | 75 | 1 | 1.8 | 2327 | 52 | 41.4% | 235526 | | |
| L3 | 3.8 | 75 | 1 | 1.8 | 76 | 144 | 114.7% | 7692 | 7844 | 397 |
| L3 | 3.8 | 75 | 1 | 1.8 | 79 | 136 | 108.4% | 7996 | | |
| L4 | 3.8 | 75 | 1 | 1.8 | 1599 | 35 | 27.9% | 161842 | 158350 | 8023 |
| L4 | 3.8 | 75 | 1 | 1.8 | 1530 | 33 | 26.3% | 154858 | | |
| L5 | 3.8 | 75 | 1 | 1.8 | 113 | 140 | 111.6% | 11437 | 11184 | 567 |
| L5 | 3.8 | 75 | 1 | 1.8 | 108 | 144 | 114.7% | 10931 | | |
| L6 | 3.8 | 75 | 1 | 1.8 | 2327 | 50 | 39.8% | 235526 | 235526 | 11933 |
| L6 | 3.8 | 75 | 1 | 1.8 | 2327 | 47 | 37.5% | 235526 | | |
| L7 | 3.8 | 75 | 1 | 1.8 | 78 | 128 | 102.0% | 7895 | 9413 | 477 |
| L7 | 3.8 | 75 | 1 | 1.8 | 108 | 128 | 102.0% | 10931 | | |
| L8 | 3.8 | 75 | 1 | 1.8 | 1665 | 30 | 23.9% | 168522 | 168117 | 8518 |
| L8 | 3.8 | 75 | 1 | 1.8 | 1657 | 23 | 18.3% | 167713 | | |
| L9 | 3.8 | 75 | 1 | 1.8 | 40 | 144 | 114.7% | 4049 | 4150 | 210 |
| L9 | 3.8 | 75 | 1 | 1.8 | 42 | 139 | 110.8% | 4251 | | |
| L10 | 3.8 | 75 | 1 | 1.8 | 2053 | 28 | 22.3% | 207794 | 210830 | 10682 |
| L10 | 3.8 | 75 | 1 | 1.8 | 2113 | 21 | 16.7% | 213866 | | |
| L11 | 3.8 | 75 | 1 | 1.8 | 119 | 126 | 100.4% | 12045 | 12500 | 633 |
| L11 | 3.8 | 75 | 1 | 1.8 | 128 | 116 | 92.4% | 12955 | | |
| L12 | 3.8 | 75 | 1 | 1.8 | 861 | 11 | 8.8% | 87146 | 88563 | 4487 |
| L12 | 3.8 | 75 | 1 | 1.8 | 889 | 18 | 14.3% | 89980 | | |
| NEC_USN | 0.69 | 75 | 1 | 1.8 | 0 | 126 | 100.4% | 0 | 0 | 0 |
| NEC_USN | 0.69 | 75 | 1 | 1.8 | 0 | 125 | 99.6% | 0 | | |
| PUC | 3.8 | 75 | 1 | 1.8 | 41 | 97 | 77.3% | 4150 | 3796 | 192 |
| PUC | 3.8 | 75 | 1 | 1.8 | 34 | 125 | 99.6% | 3441 | | |

## A-5 Proof of Principle Primer Panel

| Primer Name | Forward Sequence | Reverse Sequence | chr | Amplicon Start | Amplicon ENd |
|---|---|---|---|---|---|
| EGFR_E00001601336_1 | GCGTCTTCACCTGGAAGGG | CCGGACATAGTCCAGGAGG | 7 | 55248902 | 55249111 |
| EGFR_E00001601336_2 | GCGTGGACAACCCCCAC | GGCTCCTTATCTCCCCTCC | 7 | 55249005 | 55249213 |
| EGFR_E00001631695_1 | GTGTCACTCGTAATTAGGTCCA | GGCCTCAGTACAAACTCATTAGC | 7 | 55260366 | 55260575 |
| EGFR_E00001681524_1 | GGATGCAGAGCTTCTTCCCA | TTCTCTTCCGCACCCAG | 7 | 55259352 | 55259542 |
| EGFR_E00001681524_2 | GGTCTTCTCTGTTTCAGGGCAT | GCTGACCTAAAGCCACCTCC | 7 | 55259395 | 55259591 |
| EGFR_E00001773562_1 | TACCCTCCATGAGGCACAC | GGAGAgCTGTAAATTCTGGCTT | 7 | 55269336 | 55269516 |
| EGFR_E00001779947_1 | TGTTCATTCATGATCCCACTGC | CCACCAGTCACTCACACTTG | 7 | 55266362 | 55266571 |
| EGFR_E00001779947_2 | TCCCTGCCAGCGAGAT | AGGGATGCAAAGGCCTCA | 7 | 55266461 | 55266641 |
| EGFR_E00001790701_1 | GCCTTCTTTAAGCAATGCCATCTTTAT | CAATGGAAGCaCAGACTGCAA | 7 | 55267932 | 55268134 |
| EGFR_E00001801208_1 | CCCCTGCTCCTATAGCCAA | ATGAGGTACTCGTCGGCATC | 7 | 55268806 | 55268987 |
| EGFR_E00001801208_2 | ACTTCTACCGTGCCCTGA | GTTCAAATGAGTAGACACAGCTT | 7 | 55268921 | 55269101 |
| EGFR_Exon19 | TCACAATTGCCAGTTAACGTCT | CCACACAGCAAAGCAGAAAC | 7 | 55242373 | 55242537 |
| PTEN_E00001156315_1m | GCAACAGATAACTCAGATTGCCTT | GTTTCCTCTGGTCCTGGTATGA | 10 | 89720457 | 89720706 |
| PTEN_E00001156315_5 | AGGACAAAATGTTTCACTTTTGGGTAA | ACTAGATATTCCTTGTCATTATCTGCAC | 10 | 89720649 | 89720799 |
| PTEN_E00001156315_7 | CCTCAGAAAAAgTAGAAAATGGAAGTC | ACAAGTCAaCAACCCCCACA | 10 | 89720706 | 89720915 |
| PTEN_E00001156321_1 | TGACAGTTTGACAGTTAAAGGCAT | CACACACAGGTAACGGCTGA | 10 | 89717547 | 89717726 |
| PTEN_E00001156321_2 | TGTGGTCTGCCAGCTAAAGG | TCTCCCAATGAAAGTAAAGTACAAACC | 10 | 89717620 | 89717802 |
| PTEN_E00001156321_4m | TCCACAAACAGAACAAGATGCT | GGCCTTTTCCTTCAAACAGGATT | 10 | 89717748 | 89717956 |
| PTEN_E00001156327_1 | TCTTAAATGGCTACGACCCAG | TCCAGATGATTCTTTAACAGGTAGC | 10 | 89711775 | 89711942 |
| PTEN_E00001156327_4 | CAGTCAGAGGCGCTATGTGT | TCTAGATATGGTTAAGAAAACTGTTCCA | 10 | 89711889 | 89712077 |
| PTEN_E00001156330_1 | ttCTTATTCTGAGGTTATCTTTTTACCAC | TCATTACACCAGTTCGTCCCT | 10 | 89692739 | 89692919 |
| PTEN_E00001156330_3 | TGACCAATGGCTAAGTGAAGATGA | TCCAGGAAGAGGAAAGGAAAAACA | 10 | 89692840 | 89693048 |
| PTEN_E00001156337_4 | TATATCACTTTTAAACTTTTCTTTTAGTTGTGC | CTCGATAATCTGGATGACTCATTATTGTT | 10 | 89690776 | 89690940 |
| PTEN_E00001156344_1 | AATCTGTCTTTTGGTTTTTCTTGATAGT | AATAGTTGTTTTAGAAGATATTTGCAAGC | 10 | 89685172 | 89685367 |
| PTEN_E00001156351_1 | TGCTGCATATTTCAGATATTTCTTTCCTTA | ATGAAAACACAACATGAATATAAACATCAAT | 10 | 89653738 | 89653927 |
| PTEN_E00001456541_1 | AGATGAGTcATATTTGTGGGTTTTCA | TCTGGATCAGAGTCAGTGGT | 10 | 89724997 | 89725180 |
| PTEN_E00001456541_2 | GTAGAGGAGCCGTCAAATCCA | TTCATGGTGTTTTATCCCTCTTGA | 10 | 89725068 | 89725264 |
| PTEN_E00001456562_1 | GCAGCTTCTGCCATCTCTCT | TCCGTCTACTCCCACGTTCT | 10 | 89624175 | 89624372 |
| TP53_00001404886_13 | tctgTATCAGGCAAAGTCATAGAA | GCCTCAAAGACAATGGCTCC | 17 | 7576584 | 7576734 |

| | | | | | |
|---|---|---|---|---|---|
| TP53_E00001255919_1 | GAGAAAGCCCCCCTACTGC | AGCATCTTATCCGAGTGGAAGG | 17 | 7578091 | 7578274 |
| TP53_E00001255919_3 | TCCAAATACTCCACACGCAAA | gCTGCCCCCACCATGAG | 17 | 7578229 | 7578406 |
| TP53_E00001255919_5 | AGCTGCTCACCATCGCTA | CCAACTGgCCAAGACCT | 17 | 7578361 | 7578525 |
| TP53_E00001255919_6 | TGTGCTGTGACTGCTTGTAG | TGCCCTGACTTTCAACTCTGT | 17 | 7578425 | 7578594 |
| TP53_E00001596491_1 | TTTCGCTTCCCACAGGTCTC | CAGCCAGACTGCCTTCCG | 17 | 7579758 | 7579940 |
| TP53_E00001612188_1 | ATACGGCCAGGCATTGAAGT | CCTCCTGGCCCCTGTC | 17 | 7579260 | 7579421 |
| TP53_E00001612188_2 | GGAAACCGTAGCTGCCCTG | AAGACCCAGGTCCAGATGAA | 17 | 7579359 | 7579520 |
| TP53_E00001665758_1 | GGGGTCAGaGGCAAGCAG | CTTGGGCCTGTGTTATCTCC | 17 | 7577432 | 7577631 |
| TP53_E00001728015_1 | GGAATCCTATGGCTTTCCAACC | CCCCCTCCTCTGTTGCTG | 17 | 7573859 | 7574054 |
| TP53_E00001757276_1 | GACCCAAAACCCAAAATGGC | TCCCTGCTTCTGTCTCCTAC | 17 | 7572850 | 7573030 |
| TP53_E00001789298_1 | AGAAAACGGCATTTTGAGTGT | AAGGGTGCAGTTATGCCTCA | 17 | 7576786 | 7576983 |
| TP53_E00001789298_2 | CTGGTGtTGTTGGGCAGT | ATCTCCgCAAGAAAGGGGAG | 17 | 7576908 | 7577075 |
| TP53_E00001789298_3 | TGTCCTGCTTGCTTACCTCG | GCCTCTTGCTTCTCTTTTCCT | 17 | 7577003 | 7577187 |
| KRAS_AA_HS | GCCTGCTGAAAATGACTGAA | AGAATGGTCCTGCACCAGTAA | 12 | 25398163 | 25398329 |
| BRAF_AA_HS | TCATAATGCTTGCTCTGATAGGA | CTGATGGGACCCACTCCAT | 7 | 140453108 | 140453256 |
| EXP0116_TP53_E4 | CAGCCTCTGGcATTCTGG | cctggtcctctgaCTGCTCT | 17 | 7579479 | 7579626 |
| EXP0116_TP53_E3 | TCAAATCATCCATTGCTTGG | ccatgggactgactttctgc | 17 | 7579557 | 7579754 |
| PIK3CA_Ex10_gap | cAGAGGGGAAAAATaTGACAAA | AACAGAGAATCTCCATTTTAGCAC | 3 | 178935943 | 178936150 |
| PIK3CA_Exn21_gap | TGAGCAAGAGGCTTTGGAGT | GGTCTTTGCCTGCTGAGAGT | 3 | 178952038 | 178952227 |

## A-6 *TP53* Primer Panel

| Primer name | Forward sequence | Reverse sequence | Chromosome | amp start | amp end |
|---|---|---|---|---|---|
| TP53_D0008_001F | GTGGGGAACAAGAAGTGGAGA | CTCCCTGCTTCTGTCTCCTAC | 17 | 7572903 | 7573031 |
| TP53_D0008_002F | AGGAAGGGGCTGAGGTCA | GCGCTTCGAGATGTTCCGA | 17 | 7573904 | 7574019 |
| TP53_D0008_003F | TTGAGTTCCAAGGCCTCATTCA | GTACTGTGTATATACTTACTTCTCCCCC | 17 | 7573975 | 7574077 |
| TP53_D0008_004F | AAACGGCATTTTGAGTGTTAGAC | ACAACACCAGCTCCTCTCC | 17 | 7576789 | 7576917 |
| TP53_D0008_005F | CATCCAGTGGTTTCTTCTTTGG | TTCACTTTTATCACCTTTCCTTGCC | 17 | 7576873 | 7576961 |
| TP53_D0008_006F | CGCTTCTTGTCCTGCTTGC | TGTGCCTGTCCTGGGAGA | 17 | 7576996 | 7577115 |
| TP53_D0008_007F | ATTCTCTTCCTCTGTGCGCC | TTGCTTCTCTTTTCCTATCCTGAGT | 17 | 7577074 | 7577182 |
| TP53_D0008_008F | GGTCAGAGGCAAGCAGAGG | CCTCACCATCATCACACTGGAA | 17 | 7577434 | 7577528 |
| TP53_D0008_009F | CAAGTGGCTCCTGACCTGG | CTAGGTTGGCTCTGACTGTACC | 17 | 7577484 | 7577612 |
| TP53_D0008_010F | ACTGTTACACATGTAGTTGTAGTGGA | CTTGCCACAGGTCTCCCC | 17 | 7577561 | 7577667 |
| TP53_D0008_011F | GGCCACTGACAACCACCC | TCGACATAGTGTGGTGGTGC | 17 | 7578120 | 7578213 |
| TP53_D0008_012F | AGACCTCAGGCGGCTCA | CCTCAGCATCTTATCCGAGTGG | 17 | 7578173 | 7578278 |
| TP53_D0008_013F | ATCCAAATACTCCACACGCAAATTT | CCTCTGATTCCTCACTGATTGCT | 17 | 7578228 | 7578317 |
| TP53_D0008_014F | TGTCGTCTCTCCAGCCCC | GGCCATCTACAAGCAGTCACA | 17 | 7578343 | 7578450 |
| TP53_D0008_015F | GCCTCACAACCTCCGTCAT | AACTGGCCAAGACCTGCC | 17 | 7578407 | 7578523 |
| TP53_D0008_016F | GGAATCAACCCACAGCTGC | AACTCTGTCTCCTTCCTCTTCCTA | 17 | 7578483 | 7578581 |
| TP53_D0008_017F | TGGAAGCCAGCCCCTCA | GTCATCTTCTGTCCCTTCCCAG | 17 | 7579284 | 7579408 |
| TP53_D0008_018F | CCGTAGCTGCCCTGGTAG | AGCTCCTACACCGGCG | 17 | 7579364 | 7579453 |
| TP53_D0008_019F | TGACAGGGGCCAGGAG | CAATGGTTCACTGAAGACCCA | 17 | 7579405 | 7579533 |
| TP53_D0008_020F | GCATTCTGGGAGCTTCATCTG | TGACTGCTCTTTTCACCCATCT | 17 | 7579488 | 7579616 |
| TP53_D0008_021F | CCCAGCCCAACCCTTGTC | GAAGCGAAAATTCCATGGGACTG | 17 | 7579677 | 7579766 |
| TP53_D0008_022F | GCCCTTCCAATGGATCCACT | CTTCCGGGTCACTGCCAT | 17 | 7579816 | 7579928 |
| TP53_D0008_023F | CTAGGATCTGACTGCGGCTC | TGGAAGTGTCTCATGCTGGATC | 17 | 7579887 | 7579981 |

## A-7 Prostate Specific Primer Panel

| Primer name | Forward sequence | Reverse sequence | Chr | Amplicon start | Amplicon end |
|---|---|---|---|---|---|
| BRAF_D0008_001 | ACTGTTCAAACTGATGGGACCC | ACTTACTACACCTCAGATATATTTCTTCATGA | 7 | 140453098 | 140453209 |
| BRAF_D0009_001 | TCACCTATTTTTACTGTGAGGTCTTCA | TCTTCATAATGCTTGCTCTGATAGGA | 7 | 140453154 | 140453259 |
| BRAF_D0009_002 | CACATTACATACTTACCATGCCACTTT | GGACTCGAGTGATGATTGGGAG | 7 | 140481360 | 140481476 |
| EGFR_D0008_001 | TCTTCGGGGAGCAGCG | GAGACACGCCCTTACCTTT | 7 | 55086955 | 55087073 |
| EGFR_D0008_002 | GGTGGCTGGTTATGTCCTCA | GCATCATAGTTAGATAAGACTGCTAAGG | 7 | 55211009 | 55211137 |
| EGFR_D0008_003 | AGCGTGTCCTCTCTCCTCC | CGGGGTTCACATCCATCTGG | 7 | 55221681 | 55221800 |
| EGFR_D0008_004 | CATGCTCTACAACCCCACCA | GATGCCTGACCAGTTAGAGGG | 7 | 55221757 | 55221885 |
| EGFR_D0008_005 | GCTATGCAAATACAATAAACTGGAAA | GGTGACTTACTGCAGCTGTTTT | 7 | 55227942 | 55228041 |
| EGFR_D0008_006 | TGCCCACTACATTGACGGC | CTGCGTACTTCCAGACCAGG | 7 | 55232996 | 55233084 |
| EGFR_D0008_007 | ATCGGCCTCTTCATGCGA | GCCCTTGTTCCTGTCGGG | 7 | 55240746 | 55240874 |
| EGFR_D0008_008 | GTCTCTGTGTTCTTGTCCCCC | GCCCAGCACTTTGATCTTTTTGAA | 7 | 55241589 | 55241709 |
| EGFR_D0008_009 | TCTCTTGAGGATCTTGAAGGAAAC | GGGACCTTACCTTATACACCGT | 7 | 55241658 | 55241746 |
| EGFR_D0008_010 | GTTAACGTCTTCCTTCTCTCTCTGT | AGATGTTGCTTCTCTTAATTCCTTGATAG | 7 | 55242385 | 55242486 |
| EGFR_D0008_011 | CCAGAAGGTGAGAAAGTTAAAATTCCC | CCACACAGCAAAGCAGAAACT | 7 | 55242427 | 55242537 |
| EGFR_D0008_012 | CCCTCCTTCTGGCCACCA | GCACACGTGGGGGTTGT | 7 | 55248931 | 55249027 |
| EGFR_D0008_013 | GCCTACGTGATGGCCAGC | CCAGGAGGCAGCCGAA | 7 | 55248989 | 55249100 |
| EGFR_D0008_014 | AGCTCATCACGCAGCTCAT | ATCTCCCTTCCCTGATTACCTTTG | 7 | 55249062 | 55249190 |
| EGFR_D0008_015 | GCTCAACTGGTGTGTGCAG | GCTATCCCAGGAGCGCAG | 7 | 55249144 | 55249249 |
| EGFR_D0008_016 | ACAGCAGGGTCTTCTCTGTTTC | CTTGACATGCTGCGGTGTTTT | 7 | 55259388 | 55259498 |
| EGFR_D0008_017 | GGCAGCCAGGAACGTACT | TCCTTACTTTGCCTCCTTCTGC | 7 | 55259456 | 55259574 |
| EGFR_D0008_018 | CTGGGTGCGGAAGAGAAAGAATA | GCTAGTGGGAAGGCAGCC | 7 | 55259526 | 55259635 |
| KRAS_D0008_001 | CAGATCTGTATTTATTTCAGTGTTACTTACCT | CAGGCTCAGGACTTAGCAAGAA | 12 | 25378518 | 25378613 |
| KRAS_D0009_001 | TTCTTGCTAAGTCCTGAGCCTG | TTAAGGACTCTGAAGATGTACCTATGG | 12 | 25378592 | 25378690 |
| KRAS_D0008_002 | ACACAAAGAAAGCCCTCCCC | AGGAAGCAAGTAGTAATTGATGGAGA | 12 | 25380216 | 25380337 |
| KRAS_D0009_002 | TATTGTTGGATCATATTCGTCCACAAAA | CTTGTGGTAGTTGGAGCTGGT | 12 | 25398211 | 25398303 |
| KRAS_D0008_003 | GAATTAGCTGTATCGTCAAGGCAC | ATTATAAGGCCTGCTGAAAATGACTG | 12 | 25398246 | 25398337 |
| NRAS_D0012_001 | AGAAAATAATGCTCCTAGTACCTGTAGAG | GGATACAGCTGGACAAGAAGAGT | 1 | 115256400 | 115256543 |

| | | | | | |
|---|---|---|---|---|---|
| NRAS_D0008_001 | GTATTGGTCTCTCATGGCACTGT | TACCCTCCACACCCCCAG | 1 | 115256498 | 115256617 |
| NRAS_D0008_002 | AAGTGGTTCTGGATTAGCTGGATT | TTCCAACAGGTTCTTGCTGGT | 1 | 115258699 | 115258807 |
| PIK3CA_D0012_002 | CTCAAGAAGCAGAAAGGGAAGAATTT | TCTTTTCTTCACGGTTGCCTACT | 3 | 178916834 | 178916947 |
| PIK3CA_D0012_001 | CTTCGGCTTTTTCAACCCTTTTTAAAA | AATATTTTAGAAAGGGACAACAGTTAAGCT | 3 | 178916887 | 178917036 |
| PIK3CA_D0012_003 | CATTTCCACAGCTACACCATATATGAAT | ATTTGACTTTACCTTATCAATGTCTCGAAT | 3 | 178921448 | 178921589 |
| PIK3CA_D0012_004 | CCCATTATTATAGAGATGATTGTTGAATTTTCC | AACAAGTTTATATTTCCCCATGCCAAT | 3 | 178927868 | 178928011 |
| PIK3CA_D0008_001 | AGGGAAAATGACAAAGAACAGC | TTTTAGCACTTACCTGTGacTCCA | 3 | 178936028 | 178936135 |
| PIK3CA_D0012_005 | GATGCTTGGCTCTGGAATGC | TTGTCCAGCCACCATGATGT | 3 | 178951957 | 178952102 |
| PIK3CA_D0008_002 | CAAGAGGCTTTGGAGTATTTCATGAA | ATGCATGCTGTTTAATTGTGTGGA | 3 | 178952042 | 178952144 |
| TP53_D0008_001 | GTGGGGAACAAGAAGTGGAGA | CTCCCTGCTTCTGTCTCCTAC | 17 | 7572903 | 7573031 |
| TP53_D0008_002 | AGGAAGGGGCTGAGGTCA | GCGCTTCGAGATGTTCCGA | 17 | 7573904 | 7574019 |
| TP53_D0008_003 | TTGAGTTCCAAGGCCTCATTCA | GTACTGTGTATATACTTACTTCTCCCCC | 17 | 7573975 | 7574077 |
| TP53_D0008_004 | AAACGGCATTTTGAGTGTTAGAC | ACAACACCAGCTCCTCTCC | 17 | 7576789 | 7576917 |
| TP53_D0008_005 | CATCCAGTGGTTTCTTCTTTGG | TTCACTTTTATCACCTTTCCTTGCC | 17 | 7576873 | 7576961 |
| TP53_D0008_006 | CGCTTCTTGTCCTGCTTGC | TGTGCCTGTCCTGGGAGA | 17 | 7576996 | 7577115 |
| TP53_D0008_007 | ATTCTCTTCCTCTGTGCGCC | TTGCTTCTCTTTTCCTATCCTGAGT | 17 | 7577074 | 7577182 |
| TP53_D0008_008 | GGTCAGAGGCAAGCAGAGG | CCTCACCATCATCACACTGGAA | 17 | 7577434 | 7577528 |
| TP53_D0008_009 | CAAGTGGCTCCTGACCTGG | CTAGGTTGGCTCTGACTGTACC | 17 | 7577484 | 7577612 |
| TP53_D0008_010 | ACTGTTACACATGTAGTTGTAGTGGA | CTTGCCACAGGTCTCCCC | 17 | 7577561 | 7577667 |
| TP53_D0008_011 | GGCCACTGACAACCACCC | TCGACATAGTGTGGTGGTGC | 17 | 7578120 | 7578213 |
| TP53_D0008_012 | AGACCTCAGGCGGCTCA | CCTCAGCATCTTATCCGAGTGG | 17 | 7578173 | 7578278 |
| TP53_D0008_013 | ATCCAAATACTCCACACGCAAATTT | CCTCTGATTCCTCACTGATTGCT | 17 | 7578228 | 7578317 |
| TP53_D0008_014 | TGTCGTCTCTCCAGCCCC | GGCCATCTACAAGCAGTCACA | 17 | 7578343 | 7578450 |
| TP53_D0008_015 | GCCTCACAACCTCCGTCAT | AACTGGCCAAGACCTGCC | 17 | 7578407 | 7578523 |
| TP53_D0008_016 | GGAATCAACCCACAGCTGC | AACTCTGTCTCCTTCCTCTTCCTA | 17 | 7578483 | 7578581 |
| TP53_D0008_017 | TGGAAGCCAGCCCCTCA | GTCATCTTCTGTCCCTTCCCAG | 17 | 7579284 | 7579408 |
| TP53_D0008_018 | CCGTAGCTGCCCTGGTAG | AGCTCCTACACCGGCG | 17 | 7579364 | 7579453 |
| TP53_D0008_019 | TGACAGGGGCCAGGAG | CAATGGTTCACTGAAGACCCA | 17 | 7579405 | 7579533 |
| TP53_D0008_020 | GCATTCTGGGAGCTTCATCTG | TGACTGCTCTTTTCACCCATCT | 17 | 7579488 | 7579616 |
| TP53_D0008_021 | CCCAGCCCAACCCTTGTC | GAAGCGAAAATTCCATGGGACTG | 17 | 7579677 | 7579766 |

| | | | | | |
|---|---|---|---|---|---|
| TP53_D0008_022 | GCCCTTCCAATGGATCCACT | CTTCCGGGTCACTGCCAT | 17 | 7579816 | 7579928 |
| TP53_D0008_023 | CTAGGATCTGACTGCGGCTC | TGGAAGTGTCTCATGCTGGATC | 17 | 7579887 | 7579981 |
| CTNNB1_D0008_001 | AGCGGCTGTTAGTCACTGG | GTATCCACATCCTCTTCCTCAGG | 3 | 41266060 | 41266179 |
| FOXA1_D0006_001 | ACGGTTTGGTTTGTGTGGTTTT | AGCCCCATCGAGCCCT | 14 | 38060479 | 38060647 |
| FOXA1_D0006_002 | TAGTACGCCGGCTCCAGG | ACTCCTTCAACCACCCGT | 14 | 38060610 | 38060808 |
| FOXA1_D0006_003 | AGGAGGACATGAGGTTGTTGATG | GCGCCTCGGAGTTGAAGA | 14 | 38060764 | 38060955 |
| FOXA1_D0006_004 | GCGCAGTTGAGGAGGCT | TGGCGCCTCTAACCCCA | 14 | 38060917 | 38061095 |
| FOXA1_D0006_005 | ATGGAGGGGCGAGTCG | TCCGGCAACATGTTCGAGAA | 14 | 38061059 | 38061241 |
| FOXA1_D0006_006 | GCTTCTGGCGGCGCAA | GCAGAACCAGCAGCGC | 14 | 38061196 | 38061365 |
| FOXA1_D0006_007 | GAGTGGCGGATGGAGTTCT | TACCCGCACGCCAAGC | 14 | 38061327 | 38061493 |
| FOXA1_D0006_008 | TGATGAGCGAGATGTACGAGTAG | AGGCGGCCTCCATGAATG | 14 | 38061451 | 38061630 |
| FOXA1_D0006_009 | ACGGGTTCATGGCGGC | CAGTAGCCGGCATGCC | 14 | 38061580 | 38061747 |
| FOXA1_D0006_010 | TGCTGTTCATGGCGCCC | TGAACTCCATGAACACCTACAT | 14 | 38061703 | 38061867 |
| FOXA1_D0006_011 | CGCTCGTAGTCATGGTGTTCA | CCTCCAGTGCCCACCA | 14 | 38061820 | 38061978 |
| FOXA1_D0006_012 | TGCCTGCCTTGCCTGG | TGGTTGTATTGGGCAGGGTG | 14 | 38064012 | 38064205 |
| SPOP_D0006_001 | ACGGAGTCTTACAACAAGCAGG | AACCCATTTCTCCACATTTCTCCT | 17 | 47677715 | 47677910 |
| SPOP_D0006_002 | CTACTTGCCTGCTTTACCCACTA | GCAGAAATTCTCATCCTGGCC | 17 | 47679133 | 47679300 |
| SPOP_D0006_003 | TTCAACTGATCTGCACTGTGGA | AGAACTATTGCTGCTCTGATGACA | 17 | 47679254 | 47679413 |
| SPOP_D0006_004 | TTTACCCACAATGCAACATAGAATCC | TCATTTACACGGGGAAGGCTC | 17 | 47684521 | 47684679 |
| SPOP_D0006_005 | AAATCATCAGCCATTTTGTCGAGG | TTTAGTTTAGGGATTTTGTTCTGTTGGAA | 17 | 47684631 | 47684793 |
| SPOP_D0006_006 | ATGAATCTGCAGCTAAAGTGGGA | CCTTATCTGCTTGTTTCTCTTTGACTT | 17 | 47685157 | 47685323 |
| SPOP_D0006_007 | CCTTATTTCACCAAAACTATAAAATTCCTCATG | AAGGTTCCTGAGTGCCGG | 17 | 47688596 | 47688762 |
| SPOP_D0006_008 | CCACAGTCCTCCTAACTCATCTG | ACAGCAGTTAGGTAAAAGTCGTCT | 17 | 47688718 | 47688905 |
| SPOP_D0006_009 | GCTAGTCTCAGCAGAATACAAGGA | TCCCCACCCCAGAGAGTC | 17 | 47696314 | 47696482 |
| SPOP_D0006_010 | GCCTTGCACAAACCTATATGCC | CAAGGGAGAAGAAACCAAAGCTATG | 17 | 47696439 | 47696621 |
| SPOP_D0006_011 | AACCAGATCAAAGCCACAACTTG | ACCTGTTACTGGTCAGCTGTC | 17 | 47696513 | 47696688 |
| SPOP_D0006_012 | TGAATTTTGCCCGAACTTCACTC | CAGTTCTATCAAAATGGATGCTTTTGAC | 17 | 47696641 | 47696803 |
| SPOP_D0006_013 | CAGTGTCCCATAAAACCAAACAAAAC | TCCTACATGTGGACCATCAATAACTTT | 17 | 47699236 | 47699411 |
| SPOP_D0006_014 | CCCATTTCCTCCCGGCAAA | ATGCAGAGGGAGGGAGGG | 17 | 47699362 | 47699527 |
| SPOP_D0006_015 | AGATGCAAGATTCACAGGCTGA | GGTGAAGAGGGAACAGAAATCTTTG | 17 | 47700059 | 47700238 |

| | | | | | |
|---|---|---|---|---|---|
| OR5L1_D0006_001 | GGAGACATGGGCAAGGAAAAC | GACCTGAATCAGTGCAATCATG | 11 | 55578937 | 55579095 |
| OR5L1_D0006_002 | TGGAGTCACGTTGTTAGCCAA | CACCATGCACCCTAGGAAG | 11 | 55579047 | 55579239 |
| OR5L1_D0006_003 | TTAACAAGGACAAAGCCATCTCC | CAGCTCCACACGCACCT | 11 | 55579199 | 55579374 |
| OR5L1_D0006_004 | TTGCTATACACAGTCACCATGTCT | CACAGTGATATCAGAGCAAGCAAG | 11 | 55579330 | 55579524 |
| OR5L1_D0006_005 | TCTGTGATCTACCTCCTGTCTT | TGGGAAGCACAGGTGGAG | 11 | 55579475 | 55579673 |
| OR5L1_D0006_006 | CAGAGGGCAGGCACAAAG | AGTTCAGCATAGGAATCACGACT | 11 | 55579634 | 55579801 |
| OR5L1_D0006_007 | AAAGTGGCCACCGTGTTCT | CCAACACTGTGCCTCTCCC | 11 | 55579756 | 55579953 |
| AR_D0006_001 | AGTAGGTGGAAGATTCAGCCAAG | CAAACTGGCGCCGGGA | X | 66764959 | 66765150 |
| AR_D0006_002 | GTGATCCAGAACCCGGGC | CATCCTCACCCTGCTGCTG | X | 66765085 | 66765271 |
| AR_D0006_003 | GCAAGAGACTAGCCCCAGG | CAGGCTCTGGGACGCA | X | 66765225 | 66765394 |
| AR_D0006_004 | AGTGCCACCCCGAGAGA | GTGCTGGCCTCGCTCA | X | 66765359 | 66765557 |
| AR_D0006_005 | TGCTCCGCTGACCTTAAAGAC | GCGTTGTCAGAAATGGTCGA | X | 66765517 | 66765695 |
| AR_D0006_006 | CTCCAAGGACAATTACTTAGGGGG | CAAAAGTGGGGCGTACATGC | X | 66765648 | 66765807 |
| AR_D0006_007 | AGGGGAACAGCTTCGGGG | CCTTCTAGCCCTTTGGTGTAAC | X | 66765765 | 66765953 |
| AR_D0006_008 | ACTGCTGAGTATTCCCCTTTCAAG | GTCGCGACTCTGGTACG | X | 66765904 | 66766077 |
| AR_D0006_009 | GAGCACTGGACGAGGCA | GTCCCCATAGCGGCACTG | X | 66766040 | 66766218 |
| AR_D0006_010 | TGGACTACGGCAGCGC | CCATACAACTGGCCTTCTTCG | X | 66766166 | 66766331 |
| AR_D0006_011 | CATCCTGGCACACTCTCTTCA | TCGCTTTCCTGGCCCG | X | 66766286 | 66766475 |
| AR_D0006_012 | TACACTCGGCCCCCTCAG | AGGCGACATTTCTGGAAGGAA | X | 66766435 | 66766633 |
| AR_D0006_013 | ACCTGAGACTTCACTTGCCTATTT | CCATAGTGACACCCAGAAGCTT | X | 66863027 | 66863199 |
| AR_D0006_014 | AGACCTGCCTGATCTGTGGA | GTTAGTGTCTCTCTCTGGAAGGTAAA | X | 66863154 | 66863315 |
| AR_D0006_015 | ACTCATTATCAGGTCTATCAACTCTTGTAT | AGAGGAAGGAGGAGGAAGAGAAA | X | 66905810 | 66906005 |
| AR_D0006_016 | AAAAGGTAGTTGCATTGTGTGTTTTT | GCTTCTGGGTTGTCTCCTCA | X | 66931176 | 66931336 |
| AR_D0006_017 | TTCCAGCACCACCAGCC | CAAGGCTGCAAAGGAGTCG | X | 66931296 | 66931461 |
| AR_D0006_018 | GCTGGACACGACAACAACC | TCCATAGGAGCGTTCACTAAATATGAT | X | 66931420 | 66931595 |
| AR_D0006_019 | CAGACTGACCACTGCCTCTG | GCGAAGTAGAGCATCCTGGA | X | 66937285 | 66937443 |
| AR_D0006_020 | TGGCGATCCTTCACCAATGT | GCTAAGCTTCACTGTCACCCC | X | 66937400 | 66937558 |
| AR_D0006_021 | TTCCTCTGTGTATCTCCTTCCC | ATTCCCTGCACTTCTAGGCAC | X | 66941651 | 66941830 |
| AR_D0006_022 | GTCAGAAAACTTGGTGCTTTGTCTA | CATGCAATGATACGATCGAGTTCC | X | 66942590 | 66942753 |
| AR_D0006_023 | TTCTTTGATGAACTTCGAATGAACTACA | GCTCTATCAGGCTGTTCTCCC | X | 66942698 | 66942876 |

| | | | | | |
|---|---|---|---|---|---|
| AR_D0006_024 | ACAAAAGGCTGAAAGACCAAAAATCA | AAGTCCACGCTCACCATGT | X | 66943428 | 66943595 |
| AR_D0006_025 | AGTTCACTTTTGACCTGCTAATCA | TGGGGTGGGGAAATAGGGT | X | 66943547 | 66943712 |
| AKT1_D0006_001 | TGTAGCCAATGAAGGTGCCAT | GCACATCTGTCCTGGCACA | 14 | 105246485 | 105246649 |
| OR5L1_D0007_008 | *TTTTTAAGTTTCTTTTCCTCCAATC* | *CATCTGATAGTCCAAGGAGAATG* | 11 | 55578901 | 55579000 |
| HRAS_D0008_001 | TGCGCATGTACTGGTCCC | CCGGAAGCAGGTGGTCATT | 11 | 533836 | 533936 |
| HRAS_D0008_002 | GGATGGTCAGCGCACTCTT | AGGTGGGGCAGGAGAC | 11 | 534259 | 534377 |
| IDH1_D0008_001 | GCCAACATGACTTACTTGATCCC | GGTGGCACGGTCTTCAGAG | 2 | 209113078 | 209113197 |
| SPOP_D0020_001 | TGGTTTCTTCTCCCTTGGCATT | CCAAAGGGTTAGATGAAGAAAGCAA | 17 | 47696605 | 47696730 |
| SPOP_D0020_002 | CCAGTAACAGGTAAAGTGACAGGTA | GTTCTATCAAAATGGATGCTTTTGACATT | 17 | 47696677 | 47696801 |
| SPOP_D0020_003 | ATTCTTCCTCACCATTTCAGTTTATCATT | TCCTACATGTGGACCATCAATAACTTT | 17 | 47699296 | 47699411 |
| SPOP_D0020_004 | CCCATTTCCTCCCGGCAAA | CCACTTCCTATTTAATTGCTTCCTGTT | 17 | 47699362 | 47699469 |
| SPOP_D0020_005 | AGATGCAAGATTCACAGGCTGA | TCCAAGTCCTCCACCTCCG | 17 | 47700059 | 47700161 |
| SPOP_D0020_006 | CGGGGCCACTCGACATTT | CTGACTTTGGAAATCTCGTTTAACCT | 17 | 47700121 | 47700209 |
| FOXA1_D0020_007 | CTCATGCACGGGTTCATGG | CATGGGTACGGCGCTGAG | 14 | 38061573 | 38061680 |
| FOXA1_D0020_008 | CATTCATGGAGGCCGCC | GGGCGCCATGAACAGCA | 14 | 38061613 | 38061719 |
| FOXA1_D0020_009 | GTACCCATGGCCGTCACG | CATGTCCTATGCCAACCCG | 14 | 38061672 | 38061794 |
| PTEN_E00001156315_1m | GCAACAGATAACTCAGATTGCCTT | GTTTCCTCTGGTCCTGGTATGA | 10 | 89720457 | 89720706 |
| PTEN_E00001156315_5 | AGGACAAAATGTTTCACTTTTGGGTAA | ACTAGATATTCCTTGTCATTATCTGCAC | 10 | 89720649 | 89720799 |
| PTEN_E00001156315_7 | CCTCAGAAAAAgTAGAAAATGGAAGTC | ACAAGTCAaCAACCCCCACA | 10 | 89720706 | 89720915 |
| PTEN_E00001156321_1 | TGACAGTTTGACAGTTAAAGGCAT | CACACACAGGTAACGGCTGA | 10 | 89717547 | 89717726 |
| PTEN_E00001156321_2 | TGTGGTCTGCCAGCTAAAGG | TCTCCCAATGAAAGTAAAGTACAAACC | 10 | 89717620 | 89717802 |
| PTEN_E00001156321_4m | TCCACAAACAGAACAAGATGCT | GGCCTTTTCCTTCAAACAGGATT | 10 | 89717748 | 89717956 |
| PTEN_E00001156327_1 | TCTTAAATGGCTACGACCCAG | TCCAGATGATTCTTTAACAGGTAGC | 10 | 89711775 | 89711942 |
| PTEN_E00001156327_4 | CAGTCAGAGGCGCTATGTGT | TCTAGATATGGTTAAGAAAACTGTTCCA | 10 | 89711889 | 89712077 |
| PTEN_E00001156330_1 | ttCTTATTCTGAGGTTATCTTTTTACCAC | TCATTACACCAGTTCGTCCCT | 10 | 89692739 | 89692919 |
| PTEN_E00001156330_3 | TGACCAATGGCTAAGTGAAGATGA | TCCAGGAAGAGGAAAGGAAAAACA | 10 | 89692840 | 89693048 |
| PTEN_E00001156337_4 | TATATCACTTTTAAACTTTTCTTTTAGTTGTGC | CTCGATAATCTGGATGACTCATTATTGTT | 10 | 89690776 | 89690940 |
| PTEN_E00001156344_1 | AATCTGTCTTTTGGTTTTTCTTGATAGT | AATAGTTGTTTTAGAAGATATTTGCAAGC | 10 | 89685172 | 89685367 |
| PTEN_E00001156351_1 | TGCTGCATATTTCAGATATTTCTTTCCTTA | ATGAAAACACAACATGAATATAAACATCAAT | 10 | 89653738 | 89653927 |
| PTEN_E00001456541_1 | AGATGAGTcATATTTGTGGGTTTTCA | TCTGGATCAGAGTCAGTGGT | 10 | 89724997 | 89725180 |

| | | | | | |
|---|---|---|---|---|---|
| PTEN_E00001456541_2 | GTAGAGGAGCCGTCAAATCCA | TTCATGGTGTTTTATCCCTCTTGA | 10 | 89725068 | 89725264 |
| PTEN_E00001456562_1 | GCAGCTTCTGCCATCTCTCT | TCCGTCTACTCCCACGTTCT | 10 | 89624175 | 89624372 |

## A-8 Summary of TP53 mutant allele fractions determined using TAm-seq

| Sample | Chr | Position | Reference | Mutation | AF (weighted average) |
|---|---|---|---|---|---|
| 002_blood | 17 | 7577565 | T | G | 0.04% |
| 002_plasma | 17 | 7577565 | T | G | 1.54% |
| 002_tissue01 | 17 | 7577565 | T | G | 76.09% |
| 002_tissue02 | 17 | 7577565 | T | G | 0.97% |
| 002_tissue03 | 17 | 7577565 | T | G | 45.29% |
| 002_tissue04 | 17 | 7577565 | T | G | 0.00% |
| 014_blood01 | 17 | 7579422 | G | A | 0.26% |
| 014_blood01 | 17 | 7579421 | G | A | 0.46% |
| 014_blood01 | 17 | 7577121 | G | A | 0.25% |
| 014_plasma | 17 | 7579422 | G | A | 0.36% |
| 014_plasma | 17 | 7579421 | G | A | 0.64% |
| 014_plasma | 17 | 7577121 | G | A | 30.00% |
| 014_tissue01 | 17 | 7579422 | G | A | 0.20% |
| 014_tissue01 | 17 | 7579421 | G | A | 0.16% |
| 014_tissue01 | 17 | 7577121 | G | A | 37.65% |
| 014_tissue02 | 17 | 7579422 | G | A | 0.78% |
| 014_tissue02 | 17 | 7577121 | G | A | 48.09% |
| 014_tissue02 | 17 | 7579421 | G | A | 1.02% |
| 014_tissue03 | 17 | 7579422 | G | A | 1.36% |
| 014_tissue03 | 17 | 7577121 | G | A | 1.66% |
| 014_tissue03 | 17 | 7579421 | G | A | 1.20% |
| 014_tissue04 | 17 | 7577121 | G | A | 35.45% |
| 014_tissue04 | 17 | 7579421 | G | A | 0.18% |
| 014_tissue04 | 17 | 7579422 | G | A | 0.20% |
| 068_blood | 17 | 7577139 | G | A | 0.18% |
| 068_bloodrep_2 | 17 | 7577139 | G | A | 0.19% |
| 068_Tissue02 | 17 | 7577139 | G | A | 32.98% |
| 068_Tissue03 | 17 | 7577139 | G | A | 0.17% |
| 068_Tissue04 | 17 | 7577139 | G | A | 0.19% |
| 068_Tissue05 | 17 | 7577139 | G | A | 0.65% |
| 094_blood | 17 | 7577085 | C | T | 0.00% |
| 094_bloodrep_2 | 17 | 7577085 | C | T | 0.15% |
| 094_tissue | 17 | 7577085 | C | T | 33.02% |
| 130_blood | 17 | 7577121 | G | A | 0.24% |
| 130_bloodrep_2 | 17 | 7577121 | G | A | 0.24% |
| 130_tissue | 17 | 7577121 | G | A | 60.96% |
| 299_blood01 | 17 | 7579313 | G | A | 0.19% |
| 299_blood01 | 17 | 7578475 | G | A | 0.30% |
| 299_blood01rep_2 | 17 | 7578475 | G | A | 0.36% |
| 299_blood01rep_2 | 17 | 7579313 | G | A | 0.20% |
| 299_plasma | 17 | 7579313 | G | A | 0.16% |

| | | | | | |
|---|---|---|---|---|---|
| 299_plasma | 17 | 7578475 | G | A | 0.12% |
| 299_tissue01 | 17 | 7579313 | G | A | 0.20% |
| 299_tissue01 | 17 | 7578475 | G | A | 0.22% |
| 299_tissue04 | 17 | 7579313 | G | A | 0.43% |
| 299_tissue04 | 17 | 7578475 | G | A | 0.38% |
| 299blood_02 | 17 | 7578475 | G | A | 0.25% |
| 299blood_02 | 17 | 7579313 | G | A | 0.28% |
| 299tissue_02 | 17 | 7579313 | G | A | 0.23% |
| 299tissue_02 | 17 | 7578475 | G | A | 0.32% |
| 299tissue_03 | 17 | 7578475 | G | A | 66.49% |
| 299tissue_03 | 17 | 7579313 | G | A | 3.53% |
| 486_blood | 17 | 7577124 | C | G | 0.00% |
| 486_bloodrep_2 | 17 | 7577124 | C | G | 0.01% |
| 486_plasma | 17 | 7577124 | C | G | 0.26% |
| 486_tissue01 | 17 | 7577124 | C | G | 0.35% |
| 486_tissue02 | 17 | 7577124 | C | G | 58.72% |
| 498_blood01 | 17 | 7578528 | A | T | 0.31% |
| 498_blood01 | 17 | 7578526 | C | T | 0.34% |
| 498_blood01 | 17 | 7578410 | T | A | 0.07% |
| 498_blood01rep_2 | 17 | 7578526 | C | T | 0.29% |
| 498_blood01rep_2 | 17 | 7578410 | T | A | 0.10% |
| 498_blood01rep_2 | 17 | 7578528 | A | T | 0.41% |
| 498_blood02 | 17 | 7578526 | C | T | 0.57% |
| 498_blood02 | 17 | 7578528 | A | T | 0.54% |
| 498_blood02 | 17 | 7578410 | T | A | 0.29% |
| 498_plasma | 17 | 7578528 | A | T | 0.06% |
| 498_plasma | 17 | 7578526 | C | T | 0.19% |
| 498_plasma | 17 | 7578410 | T | A | 0.13% |
| 498_tissue01 | 17 | 7578528 | A | T | 0.05% |
| 498_tissue01 | 17 | 7578410 | T | A | 0.89% |
| 498_tissue01 | 17 | 7578526 | C | T | 0.31% |
| 498_tissue02 | 17 | 7578526 | C | T | 24.81% |
| 498_tissue02 | 17 | 7578410 | T | A | 0.02% |
| 498_tissue02 | 17 | 7578528 | A | T | 60.45% |
| 498_tissue03 | 17 | 7578410 | T | A | 0.06% |
| 498_tissue03 | 17 | 7578528 | A | T | 65.35% |
| 498_tissue03 | 17 | 7578526 | C | T | 27.91% |
| 498_tissue04 | 17 | 7578410 | T | A | 0.08% |
| 498_tissue04 | 17 | 7578526 | C | T | 0.58% |
| 498_tissue04 | 17 | 7578528 | A | T | 0.70% |
| 513_blood | 17 | 7573952 | G | A | 0.18% |
| 513_bloodrep_2 | 17 | 7573952 | G | A | 0.17% |
| 513_tissue01 | 17 | 7573952 | G | A | 1.70% |
| 513_tissue02 | 17 | 7573952 | G | A | 0.14% |
| CamP_FFPE_11 | 17 | 7579422 | G | A | 6.65% |

## A-9 Multifocal Prostate Cancer Case 7 Primer Panel.

| Primer name | Fwd sequence | Rev sequence | Chromosome | amp start | amp end |
|---|---|---|---|---|---|
| Exon_ABI3BP_7_T2F | GTACATGACTAACGTACCTGCACTT | CTTGCCACCCAGACCTACAC | 3 | 100497111 | 100497204 |
| Exon_ADAM28_7_NF | TTTTGTTTTTCTACAGTTTAAAAAGGCTGG | AGATTTACCATTACACATTTCAGGCAG | 8 | 24192943 | 24193033 |
| Exon_ATF1_7_N+7_T1+7_T2F | GTATGTGACAAAAGATACCATAAAGAAAC | CCTTCTGTTAGGTCTCTTATTTACAG | 12 | 51211428 | 51211575 |
| Exon_AUTS2_7_T1+7_T2F | gtgatccacccacctTCTTTT | TGTACAACATCTCAGTGTCCAC | 7 | 70186522 | 70186621 |
| Exon_BCAT1_7_NF | TGTCACTCCTGGAAGAATGATGC | TGTAGTGAAAGTTTGAAAATGTCCCAA | 12 | 24989478 | 24989571 |
| Exon_C1RL_X_F | ctaaactttcagtaaatacctgccttgg | ttctccgtgtaactcaattttgcc | 12 | 23274587 | 23274677 |
| Exon_CEP112_7_NF | GGTCTTCTGAAGCTGGAGATGAA | CACGTGAGAAGACCGAATCCTC | 17 | 63640501 | 63640591 |
| Exon_CHPF2_7_NF | CCTGCTGTTGGAATGTGTGAC | GATTTCCACCCGGCTCAGT | 7 | 150934789 | 150934879 |
| Exon_CLEC12A_7_T1F | ggaggattgggttgtgagtgaa | tttcgtgacaaactttcttctccaaa | 12 | 10159148 | 10159239 |
| Exon_CXorf38_7_T1F | cttttggctatagtgaatggagc | aaGaaacaattaaacagttaccacatgac | X | 40503305 | 40503440 |
| Exon_DEPDC1B_7_T1+7_T2F | ttggtgctgcagacatttactaa | ctttctttccctgtattatggtggac | 5 | 59957684 | 59957790 |
| Exon_DLC1_7_T1F | AGGTAATCAGGTATGTACAGCCACTA | ACTCTCACTACTATAGCAAAGTCTTATAGAAAT | 8 | 13334861 | 13334970 |
| Exon_FAT2_7_NF | GTCAGAGCCCTCATAGTTGGG | CAAGGGGGTGGGTATGCG | 5 | 150885171 | 150885291 |
| Exon_GGNBP1_7_NF | gtctcaactttgctattcctttgctat | agatgcacaggaaagcttaca | 6 | 33554891 | 33555017 |
| Exon_GSDMD_7_T3F | AACTGGAGCTTTTGGACAGAGAG | GCTCACCGCCTCCTCCAA | 8 | 144644204 | 144644307 |
| Exon_KIT_7_T1+7_T2F | GAGATCCTGGATGAAACGAATGAGAA | GTTTGTTGGTGCACGTGTATTTG | 4 | 55561815 | 55561911 |
| Exon_LHFPL3_7_N+7_T1+7_T2F | cgtgcccagccGGATCTT | AGATCTATAAACTTCTGAATAGCTCCAGG | 7 | 104449630 | 104449728 |
| Exon_MIPOL1_7_T3F | ACCCTTAGTACATATGAAGAAGCTTTAAAA | GTTGCAGTTGAGTAGCCAGTTC | 14 | 37969165 | 37969258 |
| Exon_PARP3_7_T3F | TATGAGGACTACAACTGCACCC | AGCAGGTGAAGAAGCGGTT | 3 | 51978108 | 51978213 |
| Exon_POTEA_7_T3F | tcacctctcactgcatcctct | gtacaaagaaattagctgggcat | 8 | 43210586 | 43210698 |
| Exon_RFTN1_7_T5F | CACTATGCCCCTCCTCTGTG | TCTCCCTGGGTCGTTCATTAAA | 3 | 16408330 | 16408436 |
| Exon_SF3B1_7_T1+7_T2F | CATAAAGGCTTTAACACAGAATCAAAAGAT | GTTTTGTAGGTCTTGTGGATGAGC | 2 | 198266737 | 198266863 |
| Exon_SH3RF2_7_T3F | GGTGTCCTCGTCCTCCAGA | TGTTGTGATGGAAAACTGCCC | 5 | 145393384 | 145393474 |
| Exon_STAB1_7_T3F | GGTGAGCAGCGACCATGAG | CATCAGGCTCCCAGTGCAG | 3 | 52553312 | 52553403 |
| fusion_7_T1-7_T2_FWD | TAACTGTGGCCTCCACAACA | TCTGGTCTGAACGTTGCACT | 21 | 39872107 | 42861556 |
| fusion_7_T2_FWD | GTTTCCCAAGGCCACATAAG | TGATCCTCCAGTCTTGAAACA | 21 | 39831176 | 42875664 |
| fusion_7_T3_FWD | TGCTTCCAGTTTAGGAGTTCA | AGTGCTCTGTGAAGCCCTGT | 21 | 39861534 | 42865380 |
| fusion_7_T4_FWD | CATTGGATCCATAGCCTGGT | TTGTCATGAAGGAACTCCTGTTT | 21 | 39835710 | 42867198 |

| | | | | | |
|---|---|---|---|---|---|
| fusion_7_T4-7_T5_FWD | GTATAACACATTTGGATGTTTTGC | CCTATCATTACTCGATGCTGTTG | 21 | 39868687 | 42870094 |
| Inter_7_N_R30F | ATCAGAGCAGGTTTGCAGGA | TCATTGAGAGCTCCCCTTTT | 5 | 110209013 | 110209100 |
| Inter_7_N_R32F | CCTTCCATACTGCCCTAGCA | GGATGTACGGAAACATCTGGA | X | 141602686 | 141602775 |
| Inter_7_N+7_T1+7_T2_R33F | AAGTTGGAACAATCTGGAGGT | TCTGGAAGTGGAATTTCTAAGTCA | 6 | 45861989 | 45862078 |
| Inter_7_N+7_T1+7_T2_R34F | GAGGCTGGCAATTTCAAGAG | CCATGTTATGCCTTCTGCAA | 8 | 29553417 | 29553487 |
| Inter_7_T1_R35F | GAATGACCTGGTAGAGCTTGG | GCAAAAGTTTAGGAACCTCTGG | 12 | 41275171 | 41275280 |
| Inter_7_T1_R36F | CCCTACCCGGTGTAGTGCTT | TGCTGCTTTCATTCTTCCAA | 3 | 117270159 | 117270260 |
| Inter_7_T1_R37F | TGTCAAAATCAGGTGTTGTAAAGC | AGCGAGAACCTGTCTCCAAA | 5 | 166259124 | 166259235 |
| Inter_7_T1+7_T2_R38F | TGACCGATGTTGGTTCTTGT | TGGTTTTGCCTTCATTTTACG | 13 | 54324236 | 54324325 |
| Inter_7_T1+7_T2_R39F | TGCTCTTCCATAATGGTGTCA | TCAGCATGGAAAAACTAGAATGA | 4 | 116523668 | 116523758 |
| Inter_7_T1+7_T2_R41F | TCTGGCAGAGACGTCTGAAA | CCAAATTACCAGTATTTACTTCAGAAAG | 5 | 18258995 | 18259149 |
| Inter_7_T1+7_T2_R42F | GACTAAAGGTTTGGTGTGAGGA | ACTCAGGCCCTGCTGTTCT | 10 | 109614775 | 109614867 |
| Inter_7_T1+7_T2_R43F | TGGAATGATGAGAATGAAGCA | TGGGTCTGTGGTATTCCATTT | 3 | 87172178 | 87172266 |
| Inter_7_T1+7_T2_R44F | GTGCAGCAGAGAAAATCCAA | TTACTGCCCACTCCACACTG | 13 | 80622031 | 80622120 |
| Inter_7_T1+7_T2+7_T3+7_T4+7_T5_R45F | TCCCAAGTAGCTGGGACTACA | CTGGCCAACATGGTGAAAC | 1 | 5130934 | 5130934 |
| Inter_7_T2_R40F | TCCCAAAGACTGAAAATTAGGA | CCAGATGTTTGACCTTGAGAAA | 5 | 130131369 | 130131467 |
| Inter_7_T2_R46F | TCCCTGTGGAATGTTTGTGA | TGCTTGCAAAGGAATTACACC | 6 | 97811929 | 97812021 |
| Inter_7_T2_R47F | GCACAACAGTAATGACTTCCATTT | GGGCGATGTGATTGCCTAT | 11 | 41199267 | 41199359 |
| Inter_7_T2_R48F | CCTACACCAGCTAAAGAGGTCA | TGCTCCATAAAGGAGGAGAAA | 4 | 78936913 | 78937005 |
| Inter_7_T2_R49F | CATCTAGTTATTTCTAATCACTTGGTAGC | TGGCCACCTAAATTCTAGAGC | 1 | 79871422 | 79871541 |
| Inter_7_T2_R50F | CAAAACAAACAAACAAAAAGCAC | CTGTGTCAAGATTCATGGAAGG | 14 | 26540930 | 26541010 |
| Inter_7_T3_R51F | GAGAAAATCAATTAAACCAAAAGCTG | GATTTTGATAATTCGTGTCTTTCTGT | 9 | 25277874 | 25277983 |
| Inter_7_T3_R52F | AATTCATTCCCAGAGGCAAA | GGATTCAAAAGGGCCTCAGT | 6 | 135025448 | 135025534 |
| Inter_7_T3+7_T4_R53F | AGAAGCACCCAGCAAGATGT | TGGAGTTCAGTGAGGTGTGAA | 16 | 9513898 | 9513986 |
| Inter_7_T3+7_T4_R54F | AGGTAGCCAGTCAGACGTGA | TCAACTGACTTTCCTGGTTGG | 11 | 920454 | 920529 |
| Inter_7_T3+7_T4_R55F | AGGCTCTCTCCAAGAAGAAACA | CAGGGAACAAGGGAAAACAA | 7 | 109273264 | 109273350 |
| Inter_7_T3+7_T4+7_T5_R56F | GGAAATGCGCAGAAAGAAAA | TTCATGTGCCATTACCACAA | 15 | 98904543 | 98904655 |
| Inter_7_T3+7_T4+7_T5_R57F | GCCTTGGCGTCCACTCT | GCTGGCTCTGAACTTGCTCT | 6 | 74612755 | 74612848 |
| Inter_7_T4_R58F | CGGTGGCTCACACCTGTAAT | TCAGGCTCGTCTCAAACTCC | 3 | 182725221 | 182725221 |
| Inter_7_T5_R59F | TGCCATTACAAAACGATTACAAA | TTCCTTCTCCTCCCAAATGA | 10 | 36173021 | 36173114 |

| | | | | | |
|---|---|---|---|---|---|
| Inter_7_T5_R60F | GAGGACAACATACACTCCACCA | ACACAGCTGCAGCAAATGAT | 7 | 9063467 | 9063541 |
| Inter_7_T5_R61F | TGGTCTTGGGTAAGATCTGGA | GAAACTCCACTTGCTTGAGAGAA | 5 | 28556789 | 28556880 |
| Mitochon_7_10_F | TCTCACTTCAACCTCCCTCAC | GGTTTCGATGATGTGGTCTTT | M | 13438 | 13531 |
| Mitochon_7_11_F | CATCCCTACGCATCCTTTACA | TGAAGATTAGTCCGCCGTAGT | M | 7822 | 7941 |
| Mitochon_7_12_F | ACTACGGCGGACTAATCTTCA | ACTCGATTGTCAACGTCAAGG | M | 7921 | 8011 |
| Mitochon_7_13_F | CCCACAGGTCCTAAACTACCA | GTTCGCTTTGACTGGTGAAGT | M | 2772 | 2883 |
| Mitochon_7_14_F | GCCTTCTCCTCACTCTCTCAA | TGCTAAGATTTTGCGTAGCTG | M | 4925 | 5010 |
| Mitochon_7_15_F | ACCGAACGAAAAATTCTAGGC | GAAGAGTTTTATGGCGTCAGC | M | 3374 | 3475 |
| Mitochon_7_16_F | AGCCAATATTGTGCCTATTGC | AGGCCATATGTGTTGGAGATT | M | 10635 | 10730 |
| Mitochon_7_17_F | TCCCCCATTATTCCTAGAACC | TATTATACGAATGGGGGCTTC | M | 7955 | 8045 |
| Mitochon_7_18_F | GCCGCAGTACTGATCATTCTAT | TTAGTCATTGTTGGGTGGTGA | M | 8582 | 8675 |
| Mitochon_7_19_F | ATTTAGCTGACTCGCCACACT | CCACCTACGGTGAAAAGAAAG | M | 6864 | 6959 |
| Mitochon_7_20_F | AGCAGGAACAGGTTGAACAGT | ATGGCCCCTAAGATAGAGGAG | M | 6267 | 6389 |
| Mitochon_7_21_F | CAGCACCACGACCCTACTACT | AGGAGGGTGGATGGAATTAAG | M | 5139 | 5216 |
| Mitochon_7_22_F | AAAATTCTCCGTGCCACCTAT | CGGGAAGGGTATAACCAACAT | M | 4367 | 4465 |
| Mitochon_7_23_F | GCTCTCCATGCATTTGGTATT | GATACTGCGACATAGGGTGCT | M | 38 | 128 |
| Mitochon_7_24_F | CCAACCTCCTACTCCTCATTG | GGGCCTTTGCGTAGTTGTAT | M | 3318 | 3417 |
| Mitochon_7_25_F | CCGGCGTAAAGAGTGTTTTAG | AATCCCAGTTTGGGTCTTAGC | M | 933 | 1079 |
| Mitochon_7_26_F | AAAAGAGCACACCCGTCTATG | CTTGGACAACCAGCTATCACC | M | 1937 | 2030 |

# A-10 Multifocal Prostate Cancer Case 8 Primer Panel.

| Primer name | Fwd sequence | Rev sequence | Chromosome | amp start | amp end |
|---|---|---|---|---|---|
| Exon_AASDH_8_T1F | TATCCCTGTAGTCCCTGATGTATGTA | AAATAAAAAGCATAAGTTCTGAGCATGTC | 4 | 57244370 | 57244485 |
| Exon_ATM_8_T1+8_T2F | GAAAATTCTTCTTGCCATATGTGAGC | CTCCTAAGCCACTTTTTATATCTTTCAGTAAT | 11 | 108160340 | 108160467 |
| Exon_CDK12_8_T3F | gttgggaatacaggcgcat | aggctgaggtgggtggat | 17 | 37662598 | 37662725 |
| Exon_CEACAM1_8_T1+8_T2F | GCTGAAGTTGGTTGTGTGGG | AACTGATCATTATTTGTGCTTTTCAGATG | 19 | 43013316 | 43013406 |
| Exon_CES1_8_T1+8_T2F | AGTTTACAGGGTTTGGGCTACG | CCACCACCTCTGCTGTCATG | 16 | 55853409 | 55853523 |
| Exon_GMDS_8_T2F | GCCATGTGTTTGCTGCTGT | TGTGATTTTCCAAGTGATTTCTCTT | 6 | 2087092 | 2087198 |
| Exon_PCNXL2_8_NF | tctgacttacaatggttcaacttatg | acagctgggcaaaatcatcaaatac | 1 | 233367941 | 233368088 |
| Exon_RGS7_8_T1+8_T2F | CTAGTGAGTTACAACCAATAAGTATTAAGA | ACTCAAACAGACTTTTCTCACTTTTCTAAT | 1 | 241492530 | 241492640 |
| Exon_SDC2_8_N+8_T1+8_T2+8_T3F | CTCTTTTCAGTAGGACAGTACAAAACAG | gcacatagtaaacaggcagtatatcag | 8 | 97563548 | 97563638 |
| Exon_SDK1_8_T1+8_T2F | ttgggggaaggagggagtg | atcatccatcttcagcacatct | 7 | 3686577 | 3686658 |
| Exon_SREBF1_8_T1+8_T2F | TCGCAGAGCACAGTCCTG | CACCTGGACCTGGCTTGTA | 17 | 17719841 | 17719960 |
| Exon_ZMYM4_8_T1F | CCATTTGCTCTTGGTTATTTCTTCAGT | CCAATGACATTATGCCAGTAATAGAGG | 1 | 35830502 | 35830595 |
| Inter_8_N+8_T1+8_T2+8_T3_R62F | CATTGGGAGTTGGGATTTCA | GATGACTCACCGTATCCCAGA | 3 | 151360274 | 151360374 |
| Inter_8_N+8_T1+8_T2+8_T3_R63F | TCGTATAGTTCATAGAAAATCTGAGCA | CTGGGCGACAGAGAGAGACT | 10 | 20754877 | 20754962 |
| Inter_8_N+8_T1+8_T2+8_T3_R64F | GCGCCACATTTATCCTTCTC | CTTCAGGTTTGTTCACGTTGAG | 5 | 29895550 | 29895638 |
| Inter_8_T1_R65F | GATTGGATAAAGTTGGGGTCAA | CATTCTGGATAAGACAAGAAGCTG | 9 | 30013330 | 30013417 |
| Inter_8_T1_R66F | GGGGCACGTGTTACTGATTT | CCTGTTCCACGAAATCCTCT | 1 | 235254174 | 235254240 |
| Inter_8_T1_R67F | GGAGTCCATAAGGAAGTCAGGA | GGTCATGTGGCCTTCTTCTG | 8 | 41015446 | 41015532 |
| Inter_8_T1+8_T2_R68F | TGTTCAGTTTCTTCTGTAGGGATTC | TCCACAACTGGATTGGAGAA | 12 | 80160118 | 80160223 |
| Inter_8_T1+8_T2_R69F | TCTGTGATCTCACATTCTCTTTCA | CTGGTACATACAAGGCCCTCA | 9 | 25298709 | 25298790 |
| Inter_8_T1+8_T2_R70F | TGTTTGGAATGATTATCTCTCTGA | CAGAGGTCATAAATGTAGTAAGAAACAA | 5 | 120673859 | 120673948 |
| Inter_8_T1+8_T2_R71F | TGATGGGTTTTCTGAGGTTCA | AATGACTGCCTTTCCCCTTT | 15 | 98499694 | 98499789 |
| Inter_8_T1+8_T2+8_T3_R73F | GGAGGGATGTCAGCAAGATG | CATCCATGGGTAGTTACCAAAA | 6 | 82708900 | 82708985 |
| Inter_8_T1+8_T2+8_T3_R74F | GGGAAGCAACAGATGGAAAC | TGTGTTTTCTTTCTTGACAACTCAG | 11 | 106003670 | 106003797 |
| Inter_8_T1+8_T2+8_T3_R75F | TCTCATTGTGGTTTTGATTTGC | ATACCTTACATGCGGCCAAG | 9 | 103685135 | 103685218 |
| Inter_8_T1+8_T2+8_T3_R76F | CGATATTGACAGATGTCCCTTTT | CCTCCCACCTAACCTTCTCC | 2 | 199067660 | 199067749 |
| Inter_8_T1+8_T2+8_T3_R77F | ATGCATCAAAGCTGCCTTTC | TTCCAACGAAAGGTCGTAGC | 4 | 126695839 | 126695909 |
| Inter_8_T2_R78F | CCAGCCATGTGAAACTGTGA | CATGCTGCTATGAAGAACTACCA | 3 | 190200667 | 190200747 |

| | | | | | |
|---|---|---|---|---|---|
| Inter_8_T2_R79F | TGTGGAATAGTTGCATGAAAAA | TGCTGAGATTACAGGCGTGA | 7 | 84528423 | 84528423 |
| Inter_8_T2_R80F | AACTGGGACCATAGGCATGT | CCCTGGCAATGTAGCAAGA | 1 | 120884815 | 120884902 |
| Inter_8_T2_R82F | GCAGGAACAAAAGGAAGGAA | AGTCCAAGGTCAAACAGAGCA | X | 145838513 | 145838594 |
| Inter_8_T2_R83F | TCACACCTGTAATCCCAGCA | TATATTGCCAGGGCTGGTCT | 16 | 18331328 | 18331328 |
| Inter_8_T3_R84F | TTCTGCTTCATTCACACCTTT | TCTCTCCCTCACAACATCCA | X | 65262099 | 65262173 |
| Inter_8_T3_R85F | TGCAGGTCTGTCCTTGTTTG | AATATGGGATAGGGCGAGGT | 14 | 21321829 | 21321899 |
| Inter_8_T3_R86F | TTCTTTCCCTTGGCCTCTAA | CCATGAAAGTTGCAGGAGGT | 8 | 34811291 | 34811391 |
| Inter_8_T3_R87F | CCCTAGGAATCTGCCACTGA | TCAAGTTTAGAGATTGGGAAAGAAA | 13 | 38952938 | 38953026 |
| Inter_8_T3_R88F | ATGTCTCCCAGCCTGGACT | CCCTCTTCCTGCTCTTCAGAC | 2 | 45859175 | 45859262 |
| Inter_8_T3_R89F | TCTGCAAGTCTCAGTTCCTTCA | GAACTTAGACAAAGGGATAAATTTGAG | X | 126134012 | 126134098 |
| fusion_8_T1-8_T2_FWD | GCCATTTTAATCACCCCCTAA | TTCCCGATGGATGTTTGAGT | 21 | 39831486 | 42870455 |
| Mitochon_8_27_F | CATCCCTACGCATCCTTTACA | TGAAGATTAGTCCGCCGTAGT | M | 7822 | 7941 |
| Exon_C1RL_X_F | ctaaactttcagtaaatacctgccttgg | ttctccgtgtaactcaattttgcc | 12 | 23274587 | 23274677 |